

Biodefence takes its toll

It is ironic, constitutionally questionable and misguided that in pursuit of vaccines against biowarfare agents, the Bush administration has attacked the very biomedical research budgets that have helped to make such defence possible.

Since his appointment as US secretary of health and human services, Tommy Thompson has tried to tighten his department's grip over the National Institutes of Health (NIH) in myriad ways. Some may seem petty — after all, where is the harm in asking every NIH communication to carry his department's logo? But the Bush administration's latest command to the NIH is deadly serious. It is a prime example of how centralized control could undercut the NIH's mission of protecting the health of the United States and the world.

It all started a month ago, when the White House budget office told the National Institute of Allergy and Infectious Diseases (NIAID) to spend \$233 million on purchasing a new anthrax vaccine. The administration first asked for the money in its annual budget request to Congress last year, but researchers and their advocates convinced Congress that the NIAID should not spend research dollars buying vaccines. The administration's request was denied, and Congress redistributed the money among the other NIH institutes, without specifying who would pay for the vaccine.

Now the administration has stipulated that the anthrax vaccine money must come out of the NIAID's budget — a questionable move, as Congress is supposed to have the final say over federal spending. And because the administration has demanded that the money does not come from biodefence programmes, grants in such areas as basic immunology, infectious diseases and AIDS-vaccine research are being slashed or frozen. The NIH's congressional allies have tried to intervene on its behalf, and the White House is still talking to both Congress and the NIH. But the decision does not seem likely to be reversed.

The Bush administration does not need to be reminded of the crucial importance of the worldwide fight against AIDS. On 27 May, President George W. Bush signed a law authorizing Congress to spend

\$15 billion over five years on international AIDS-relief projects. President Bush himself proposed the measure, and his administration's support helped the AIDS bill move through Congress in just four months — practically the speed of light for Washington. But it makes no sense to enact such measures while crippling the researchers who are looking for the drugs and vaccines that make such programmes possible. And while AIDS research may seem unrelated to bioterrorism, the vast majority of antiviral drugs on the market were developed as countermeasures against HIV.

The latest row over the anthrax vaccine may not seem to be closely linked to the debate about the AIDS bill. But both can be seen as cases of political manoeuvring cloaked as protection of public health. Some have criticized the AIDS bill because it requires spending on religious programmes and abstinence, which will prevent it from funding some of the key groups fighting the epidemic in developing countries. Furthermore, there is doubt over whether President Bush will ask Congress to spend the money he asked for in his bill, or whether the authorizing measure is just an empty promise. Developing a new anthrax vaccine is a worthy goal. But gutting the basic research that has yielded the best defences against anthrax and other diseases is a poor way of achieving it. In both cases, the administration is giving the impression that it cares more about its public image than it does about tackling serious health problems.

The fight to control the NIH is more than just bureaucratic squabbling. It reflects an acute concern that central command will endanger the public health that the administration should protect, and the biomedical research that is necessary for that protection. If the Bush administration's main commitment to improving the US public-health system is to double the NIH's budget, the money should be used to strengthen — not debilitate — the agency and the researchers that are best placed to help.

A meeting for Europe's scientists and publics

Scientific organizations should support the Euroscience Open Forum in 2004.

Most *Nature* readers will be aware of the annual meeting of the American Association for the Advancement of Science. It gives researchers the chance to describe their science to a diverse, non-specialized audience, offers forums for debates about issues relating to science, and provides an annual circus for hundreds of journalists from all over the globe.

In Europe, science-related issues are often more politically sensitive than in the United States, the public is fascinated by or critical of research, and there is an array of scientifically engaged media. Europe deserves such a meeting too, even if no one would expect it, in its early days, to achieve the scale and global media impact of its US equivalent.

This is why *Nature* is supporting the first Euroscience Open Forum (ESOF2004), to be held in Stockholm on 25–28 August next year (see www.esof2004.org). Stockholm's conference facilities are fully worthy of such a meeting, and sponsorship by several bodies has allowed an organizational framework to be established.

Despite the enthusiasm of its originators, the forum was not set up to pursue some European ideal. What matters is that the various themes of the meeting that are currently planned — ageing, the mind and behaviour, evolution, global change and more — are full of interesting content, and that scientists, stakeholders in science and the broader public can come together to discuss important issues, or simply to hear about hot science. *Nature* expects to support at least one such gathering at ESOF2004, and several significant European organizations are committed, or are about to commit, to partnership with the event as a whole.

Such an embryonic organization, as determined and self-motivated as it is, can benefit greatly from the support of the European Commission but must remain independent. It seems to be achieving this balance. Many laboratories and agencies are still considering their possible contributions to the event. Now is the right time for them to commit to this new conjunction of science and citizens.





Cash crisis
US museums
threatened by
economic downturn
p575



Pet problem
Could cats play a
leading role in the
fight against SARS?
p576



Deep thoughts
Commission sets out
blueprint for US
ocean policy
p577



Clean sweep
US to help replace
Russia's nuclear
facilities
p578

Harmful potential of viral vectors fuels doubts over gene therapy

Erika Check, Washington

The troubled field of gene therapy was dealt a fresh blow this week, after a study suggested that modified viruses used in some trials might cause health problems.

The study, led by geneticist Mark Kay at Stanford University, California, examined a modified virus used in gene-therapy trials to treat haemophilia and cystic fibrosis. It revealed that the virus has the potential to cause the same problems that led to cancer in an unrelated gene-therapy trial last year.

In gene therapy, doctors use a gutted virus as a 'vector' to transfer corrective genes into a patient's cells. But if the vector stitches itself into a cell's genes, it can cause the cell to mutate and become cancerous. This was demonstrated last year, when two children who had gene therapy for severe combined immunodeficiency disease (SCID) developed leukaemia (see *Nature* **419**, 545–546; 2002).

Scientists are still trying to establish exactly why the SCID patients developed cancer, and will discuss the trial at this week's meeting of the American Society of Gene Therapy in Washington DC. But most agree that gene therapy was the cause.

Kay's study focused on a vector made from an adeno-associated virus — an organism that does not cause disease in people, but



Doctors at Stanford treat a haemophiliac in a gene-therapy trial — but how safe is the procedure?

which can be engineered to infect human cells. In a paper published online on 1 June (H. Nakai *et al.* *Nature Genet.* doi:10.1038/ng1179; 2003), Kay and his colleagues show that the vector used in the haemophilia and

cystic fibrosis trials integrates itself more often into genes than it does into regions of DNA that do not contain genes. The finding suggests that the vector could potentially cause the sort of cellular defects that led

Protests win reprieve for renowned medical lab

Jim Giles, London

The UK Medical Research Council (MRC) has shelved plans to close its largest laboratory, at Mill Hill in north London — at least for the time being.

In April, the MRC proposed that the prestigious National Institute for Medical Research should move to Addenbrooke's Hospital in Cambridge, 50 miles away (see *Nature* **422**, 545; 2003). MRC officials said that the move would save money and improve interactions between basic and clinical researchers.

But the plan infuriated biomedical researchers at Mill Hill, who said they

already had a strong track record of collaborating with clinical researchers at various London hospitals, which employ a wider range of specialists than is found at Addenbrooke's.

Now their protests, which were supported by several Members of Parliament, seem to have hit home. In a letter sent on 30 May to John Skehel, a virologist and director of the Mill Hill institute, MRC chief executive George Radda said that the council would abandon its original intention of reaching a decision on the move by next month.

Radda said that the future of the institute

would instead be decided by a task force chaired jointly by himself and Colin Blakemore, the University of Oxford neuroscientist who will succeed him as chief executive in October (see *Nature* **423**, 211; 2003).

Unlike the subcommittee behind the original proposal, the task force will include scientists from Mill Hill itself, as well as from outside Britain. "This is a step forward in terms of consultation," says Skehel.

Despite their relief at the reprieve, researchers at Mill Hill say that they are still worried that continuing uncertainty could damage the institute.

to cancer in the SCID patients.

But researchers caution that the vector used in the SCID trials, which was based on a retrovirus, is very different from the adeno-associated vector. For instance, retroviruses must insert themselves into human DNA to work, but adeno-associated viral vectors integrate themselves into the genome much less often.

"Adeno-associated vectors clearly have a better safety profile than retroviral vectors," says David Russell, a geneticist at the University of Washington in Seattle. "But we really can't say yet that adeno-associated vectors won't cause cancer."

Kay's team, which is running a gene-therapy trial for haemophilia, tracked the adeno-associated viral vectors in mice. They extracted liver cells whose DNA contained the vector and then sequenced the DNA around the vector. They then analysed the sequences to see whether they matched a known gene. The team found that 72% of the time, the vector had interrupted a gene. Had it inserted itself randomly, the vector would have interrupted a gene no more than 40% of the time.

Last August, Frederic Bushman and his colleagues at the Salk Institute for Biological Sciences in La Jolla, California, suggested that retroviruses also insert themselves into genes more often than into other regions of DNA (A. R. Schroder *et al. Cell* **110**, 521–529; 2002).

Such results are leading researchers to seek better ways to target vectors to specific regions of DNA, and to develop vectors that don't integrate into DNA at all. But in the meantime, Kay says that he has taken numerous precautions to protect the 14 haemophiliacs he has treated.

"I don't think we need to modify anything at this point," Kay says. "But this is a risk we'll have to address before the vector is in widespread use."

Divisions sink US consensus effort on transgenic food

Jonathan Knight, San Francisco

An ambitious effort to hammer out an agreement between proponents and critics of agricultural biotechnology in the United States has ended in failure.

Two years of talks, convened in 2001 by the Washington-based Pew Initiative on Food and Biotechnology, reached an impasse late last month over the question of whether to ask Congress for legislation to strengthen regulations for transgenic food.

The 18-member panel was set up to produce detailed guidelines for safety standards for genetically modified products. But the ground rules for the panel required unanimous agreement on all issues under consideration before any conclusions would be made public. "It was an extremely ambitious effort," says Margaret Mellon, a panel member and director of the food and environment programme at the Union of Concerned Scientists in Washington DC. "We were going for the brass ring, and we didn't get it."

One issue that was contentious from the start was whether new laws are needed to guarantee the safety of transgenic food. The biotechnology industry opposes the involvement of Congress in favour of minor administrative changes to the current approval process. But consumer and environmental groups say that the present voluntary system is inadequate and needs a legislative fix. Panel members confirm that disagreement on this point was critical to the failure of the process.

Some observers suggest that the recent trade complaint filed at the World Trade Organization (WTO) by the United States against the European Union (see *Nature* **423**, 369; 2003) might have contributed to the



Hard to swallow: regulations on transgenic food proved a sticking point in US round-table talks.

problem by entrenching industry's position. "The case is being made on the grounds that the current regulatory process is safe," says Robert Paarlberg, an expert in international agricultural policy at Wellesley College near Boston. "The industry doesn't want any suggestion that it isn't."

Panel member Richard Caplan, an environmental advocate with the US Public Interest Research Group in Washington DC, says that although the WTO suit was not discussed in the sessions, it could have been a factor. But Mike Rodemeyer, executive director of the Pew Initiative, says that he strongly doubts whether it made any difference.

The panel's final report, issued on 30 May, left open the possibility of reconvening in 12–18 months' time. If consumers outside the United States continue to resist transgenic food, US food exporters might feel pressure to conform to tougher regulation, some panel members say.

UK universities face star treatment in funding revamp

Jim Giles, London

The seven-point scale used to assess research in British university departments could be scrapped under plans unveiled on 28 May.

The Research Assessment Exercise (RAE), a five-yearly audit of research, is widely credited with strengthening British science. But scientists have complained that the rankings are sometimes arbitrary, and give no credit to good researchers working in average departments.

Under the proposal, drawn up by Gareth Roberts, president of Wolfson College at the University of Oxford, each researcher's work would be allocated between zero and three stars, and each department's 'score' would

be the sum of the scores of its individual researchers, providing a measure of the overall volume of quality work produced.

Roberts says that individual ratings would not be published, but sceptics doubt whether the scores could be kept from researchers. "How will that information be denied them?" asks Keith Peters, an immunologist at the University of Cambridge and president of the Academy of Medical Sciences. Roberts concedes that individual rankings should not be used if they cannot be kept secret.

Roberts' review, which was commissioned by the UK higher-education funding councils, also suggests separating out the least research-intensive universities and subjecting them to a

shorter review process. That idea has drawn fire from some of the smaller universities. "It is divisive," says Mike Saks, pro-vice-chancellor for research at the University of Lincoln. The university received just £250,000 this year from the funding councils, but Saks says it has begun a major drive to improve its research and doesn't want to be put in a second class of institutions.

The funding councils, which distribute £1.1 billion (US\$1.8 billion) to universities each year, largely on the basis of the RAE, are accepting comments on the new plan until September. They will then decide how to do the next assessment, most likely in 2007.

♦ www.rareview.ac.uk



Troubled waters: San Diego's Natural History Museum, like many others, faces a budget crisis.

Natural history collections in crisis as funding is slashed

Rex Dalton, San Diego

Natural history museums and valuable university collections of plant and animal specimens are being hit hard by the economic climate across the United States, researchers say, threatening a wide variety of research projects.

Curators are being laid off, departments closed, systematic collections dispersed and library subscriptions sliced as administrators seek to make up budgetary shortfalls — often caused by cuts in support from state governments to universities and their museums. Reduced private donations, endowments hit by the stock-market slide and a downturn in revenue from visitors have also hurt.

And programmes are set to suffer more, as state governments try to balance their annual budgets for fiscal years that start in July.

"These collections help us understand the past, and be better stewards of the future," says John Heyning, deputy director for research and collections at the National History Museum of Los Angeles County, and president of the Washington-based National Science Collections Alliance. "But we've done a poor job of speaking in a strong, united voice on our role," he says — leaving collections politically vulnerable in the current climate.

Biologists say that the timing of the crisis is particularly unfortunate because researchers are just becoming able to sequence complete genomes inexpensively — potentially redefining the study of evolutionary biology and of relationships between species.

"Our collection has extraordinary untapped scientific value," says ecologist Michael Rosenzweig, director of the Museum of Natural History at Virginia Tech in Blacksburg. "But we definitely are moving backwards." State budget cuts have forced Virginia Tech to dissolve the museum's 150-year-old collection, he says, dispersing the

1.5 million specimens to various departments and institutions.

At the University of Iowa in Iowa City, budget cuts are prompting the closure of the herbarium, part of which is to be sent to Iowa State University in Ames. "This is our tool for teaching students about the environment," complains botanist Diana Horton, the herbarium's director. Its 250,000 specimens "represent the diversity" of our state, she says, noting that its intensive agriculture makes Iowa "the most altered state in the union". Horton describes the decision to get rid of a herbarium that only costs \$25,000 to maintain each year as "a joke" that is being perpetrated "to give the appearance of saving money".

The University of Nebraska, meanwhile, is looking to cut at least \$1.1 million next year from the \$1.8-million budget of the State Museum in Lincoln, which it runs for the state government. The museum holds 14 million specimens, including a scarab beetle collection on behalf of the Smithsonian Institution. Four out of nine tenured faculty positions at the museum will be eliminated, says entomologist Brett Ratcliffe, its associate research director. Four other tenured professors will be transferred to university departments, he adds.

At the San Diego Natural History Museum, scientists are unable to make full use of a new \$29-million extension because the museum is running a \$1.1-million deficit on its \$7.5-million budget for this. Nine positions — including key research staff slots — remain vacant, and nine other staff have been laid off.

The future looks bleak to custodians of natural history collections across the country, with most funding bodies in equally dire straits. The treatment of the field is "a slap in the face of the scientific community," says palaeontologist David Gillette, who lost his job at the Museum of Northern Arizona in Flagstaff only last week. ■

Biologist gets minimum sentence in 'espionage' case

David Cyranoski, Tokyo

A Japanese biologist arrested two years ago under the US economic espionage act emerged with a light sentence from a courthouse in Cleveland, Ohio, on 28 May.

But the trial has left the researcher's career in ruins — and raised worries about the use of this law against unwitting violators, and its potentially adverse effect on scientific collaboration.

Hiroaki Serizawa, then at the University of Kansas Medical Center, was arrested in May 2001 and charged under the 1996 act with allegedly helping a friend, Takashi Okamoto, by storing DNA constructs and cell lines that Okamoto had taken from his employer, the Cleveland Clinic Foundation (see *Nature* 411, 225–226; 2001).

Last year the charges against Serizawa were reduced to making "false statements" when questioned by FBI agents in 1999 (see *Nature* 417, 108; 2002). Serizawa corrected the statements on the same day he made them, the prosecutors conceded.

Serizawa was fined the minimum \$500, put on probation for three years, during which his movements within and outside the United States are restricted, and given 150 hours of community service.

This case was the first time that section 1831 of the act — in which the benefactor of the crime is a foreign entity — was invoked, and many scientists and lawyers in Japan and elsewhere questioned its use against Serizawa. Critics of the prosecution view the light sentence — as opposed to dropping charges — as an attempt to save face.

But Serizawa's career prospects remain bleak. Since his arrest in May 2001, his graduate students and assistants have left his laboratory, his university refused him tenure, and the American Cancer Society has terminated his grant. "My career was destroyed by the case," Serizawa told *Nature*.

He says he would like to continue research, but it will be difficult to find an employer. The National Institutes of Health, for example, will not give grants to people convicted of criminal offences within the past three years, unless they have special dispensation.

The United States is still seeking to extradite Okamoto from Japan. Serizawa claims Okamoto deceived him, and has filed a civil suit in Japan against Okamoto to seek compensation, including \$450,000 for legal fees. ■

Geneticists play the numbers game in vain

Helen Pearson, New York

The precise number of human genes might never be tallied, geneticists confessed this week, after handing out a cash prize for the nearest guesstimate.

To non-geneticists at least, the admission may come as a surprise. Many had assumed that the human gene count was virtually settled at between 30,000 and 40,000, an estimate announced with great fanfare in February 2001 based on early analyses of the draft human genome sequence.

But two years on, geneticists who gathered at a Cold Spring Harbor Laboratory meeting in New York state acknowledged that they are no closer to a final tally. "I'd say we don't know the true gene number for any organism — and certainly not for humans," said bioinformatics expert Phil Green of the University of Washington, Seattle.

This uncertainty has now been highlighted by a sweepstake on the human gene tally, dubbed Genesweep. The rules of the wager, which was light-heartedly set up at a Cold Spring Harbor conference in 2000, stated that the winner would be announced this year. More than 460 bets have since been placed.

On 30 May, the winning estimate — the lowest one — was announced by organizer Ewan Birney of the European Bioinformatics Institute in Hinxton, UK. The prize goes to Lee Rowen, who directs a sequencing project

at the Institute for Systems Biology in Seattle, Washington.

Rowen's wager of 25,947 is closest to the current reckoning of 24,847 made by the genetic database Ensembl. Like many good gamblers, she describes her number as "a stab"; one runner-up picked 27,462 because his date of birth was 27 April 1962.

The final gene tally is anyone's guess — but it is unlikely to rise again to the estimates of 80,000–100,000 mooted a few years ago. Geneticists at the meeting came up with many reasons why genes — regions of DNA that code for proteins — have proved so difficult to identify.

One reason is that gene-predictor programs, which trawl through DNA for landmark sequences characteristic of a gene, are notoriously unreliable. For instance, they often erroneously pick up pseudogenes, copies of real genes that have become defunct.

Conversely, the programs can miss genes carrying variations in their landmark sequences — and are completely flummoxed by unconventional cases, such as tiny genes, overlapping genes or small genes hidden within larger ones. "No gene predictor will ever get these right," says fruitfly geneticist Gerald Rubin of the University of California, Berkeley.

Lines of back-up evidence that are commonly used to strengthen program



predictions are also fallible. For example, a putative human gene is considered more likely to be real if it matches a gene also found in databases of mouse, fruitfly or other organisms. But an unknown number of human genes have no obvious match.

With so many obstacles to tracking human genes, "there will never be a final number", predicts Jean Weissenbach, director of the sequencing centre Génoscope in Evry, France. Ultimately, it may turn out that every person has a different number of genes, as mutations have eliminated minor ones from their genome, he says.

♦ www.ensembl.org/Genesweep

Pet theory comes to the fore in fight against SARS

Alison Abbott

The domestic cat is set to land a starring role in the future course of severe acute respiratory syndrome (SARS) — although it isn't yet clear whether it will be as saviour or scourge.

Experts have identified the cat as a potential animal model for SARS, but it may also fit the bill as a reservoir species for the SARS virus, guilty of bringing humans into contact with the disease.

Late last month, Chinese scientists

revealed that several species of wild animal on sale in the markets of southern China were harbouring a virus very similar to that believed to cause SARS (see *Nature* 423, 467; 2003). It is unclear whether these animals are a reservoir for the virus, or were infected by another species. And experts are now wondering whether a domestic animal may also be a reservoir species.

The only animal model available for SARS is the macaque monkey, which is expensive to use. Attempts to infect mice with the SARS virus have so far failed. But some pet cats in the Amoy Gardens apartment block in Hong Kong, where more than 100 residents contracted SARS in April (see *Nature* 423, 3–4; 2003), were found to harbour the virus, and to get sick from it.

"The cat may offer an alternative," says Albert Osterhaus, a virologist at Erasmus University in Rotterdam who developed the macaque model. Osterhaus has applied for ethical authorization to begin systematic studies on cats. Other laboratories, including the US Centers for Disease Control and Prevention in Atlanta, Georgia,

are also assessing the cat's potential as an animal model.

If cats prove to be susceptible to infection, they could be responsible for bringing SARS into homes. The World Health Organization (WHO) is keen to define all possible reservoirs among species that come into regular direct contact with humans. It is establishing an international network of laboratories with expertise in zoonotic diseases — infections that jump between species — to address the problem.

Researchers at the National Microbiology Laboratory in Winnipeg, Canada, have already tested pigs and chickens and have found no evidence that they can be infected by the virus.

"We want to recruit more specialized labs around the world to help extend this type of work to all domestic species," says Klaus Stöhr, the WHO's chief SARS expert, based at the organization's Geneva headquarters. He warns that a domestic animal acting as a reservoir would pose a greater risk to humans than if only wild species harbour the virus.



Feline dilemma: cats may prove useful as a model for SARS — and could also harbour the virus.

Panel calls for sea change to fisheries policy

Virginia Gewin, Portland

A heavyweight independent commission is calling for a drastic overhaul of US fisheries policy, and the creation of an independent government agency to manage the oceans surrounding the United States.

The Pew Oceans Commission, funded by the Pew Charitable Trusts and comprising 18 prominent environmentalists, scientists and officials from government and the fisheries industry, wants fisheries policy to be based explicitly on ecosystem management.

To prevent further harm from overfishing, ocean pollution and invasive species, the commission recommends the creation of an independent national oceans agency, the adoption of a national ocean-policy act, doubling of ocean research spending over 5 years, and the establishment of a network of national marine reserves, or protected areas.

But it remains to be seen whether the Pew commission's findings, to be released this week, will be endorsed by the government-sponsored US Commission on Ocean Policy, chaired by James Watkins, when it issues its report in the autumn. The latter is expected to shape US ocean policy for years to come (see *Nature* 418, 718; 2002).

Leon Panetta, chair of the Pew commission and former White House chief of staff, says he is confident that Watkins' commission will reinforce his own findings. "I sat down with Watkins very early on and we decided that if the two commissions came out in different places it would undermine what we are trying to do," he says. The commission chairs have since testified



Net benefit? The Pew Oceans Commission wants an independent agency to manage ocean ecosystems.

before each other and held joint meetings.

Marine experts say the Pew commission's recommendations would help to streamline the convoluted array of agencies and budgetary committees that currently manage US oceans. Representative Sam Farr (Democrat, California), co-chair of the bipartisan House Oceans Caucus, lauds the report as "a handbook for what needs to be done". But he adds that real reform proposals would have to be led by the Bush administration — and that none are expected until after Watkins' commission releases its report.

"The administration is not notable for enthusiasm for green issues, but this is a blue issue," says Charles Kennel, a member of the

Pew commission and director of the Scripps Institution of Oceanography in La Jolla, California. "The problem with the ocean is that it's nobody's backyard, it's a national resource."

The Pew commission calls for most of the ocean responsibilities of the National Oceanic and Atmospheric Administration (NOAA), as well as the ocean-related functions of other government agencies, to be transferred to a new ocean agency outside the Department of Commerce, of which NOAA is part.

But such reorganizations are notoriously hard to execute in Washington. And Bill Hogarth, deputy director of NOAA Fisheries, says that a revamp wouldn't make hard policy choices any easier.

The Pew commission's recommendation of ecosystem-based fisheries management may also prove contentious. "We haven't resolved what ecosystem-based management really means," cautions Hogarth. "Some people think it's just setting up marine protected areas."

Indeed, many marine biologists support such reserves. "If no such network of marine protected areas is implemented, there is no way that stocks aren't going to continue crashing," says Daniel Pauly, a fisheries biologist at the University of British Columbia in Vancouver, Canada.

The two commissions were established because of a widespread sense that US oceans policy was floundering. Thirty years after NOAA was created as a sort of Earthbound NASA, it has yet to match the space agency's budget or status.

Oceanographers admit that Washington's single-minded focus on terrorism makes it hard to draw attention to ocean issues. But Kennel points out that these issues aren't going to go away, and that this year's push for reform is the best chance to bring about badly needed change.

Japanese team makes stem cells

David Cyranoski, Tokyo

Japan has joined the club of countries that have produced their own human embryonic stem-cell lines.

Norio Nakatsuji and his group at Kyoto University announced on 27 May that their first stem cells had been successfully established and would be made available to other researchers in the autumn. The cells will be sold at cost to both academic and commercial researchers in Japan who have the government's permission to use them.

Under rules set in September 2001, researchers in Japan can conduct research on human embryonic stem cells, subject to approval by university and education-ministry committees. The education ministry is still deciding how to handle requests for access to the cell lines from foreign laboratories.

According to Nakatsuji, about ten

laboratories have already expressed an interest in the cell lines, and he expects 20–30 applications to use the cells by the end of the year.

His group plans to establish five more stem-cell lines for researchers. But difficulties in collecting frozen embryos — which requires extensive informed-consent procedures — have slowed down the group's progress so far, one team member says.

Several research groups in Japan have been working with human embryonic stem cells bought from suppliers abroad, such as WiCell Research Institute at the University of Wisconsin in Madison, and Monash University near Melbourne, Australia. But unlike cells from these suppliers, those from Kyoto will not come with material-transfer agreements claiming a share of the profits from any commercial products derived from the research.

International website to collate data on embryonic stem cells

London A website for assessing the merits of different embryonic stem-cell lines is to be set up by an international consortium. The plan emerged from a meeting of researchers and funders from 12 countries, held in London last week and called by Britain's Medical Research Council.

The site will contain standardized data on different cell lines, such as which sort of cell types they can differentiate into. At least one private company — ES Cell International of Singapore — has offered to share data and cell lines with the international group.

Some of the details remain to be worked out. For instance, it is not clear how the US National Institutes of Health will provide information about lines derived after 9 August 2001, as the United States has imposed restrictions on the research that can be done on cell lines created after that date. The participating countries are Australia, the United Kingdom, Israel, Germany, Singapore, Sweden, the Netherlands, Japan, Canada, France, Finland and the United States.

US set to replace Russia's last plutonium plants

Washington The last three nuclear reactors in Russia capable of producing weapons-grade plutonium are to be replaced using funding from the US Department of Energy (DOE).

The electricity-generating, uranium-burning reactors, which are located at Seversk and Zheleznogorsk in Siberia, produce enough plutonium by-product to make one nuclear weapon every 36 hours. Attempts to modify the reactors so that they would not produce weapons-grade material failed in 2000 because of technical and financial problems.

The DOE has now agreed to pay the US defence giant Raytheon Corporation US\$466 million to replace the nuclear reactors with coal-fired power plants, as part

of a plan to reduce the spread of weapons-grade nuclear material. The facilities are the last of Russia's 13 plutonium-producing reactors to be decommissioned.

Canadians see stars after cash windfall

Montreal Canadian astronomers had double cause for celebration last week — a funding boost of Can\$50 million (US\$37 million) over five years will allow them to finalize a major collaborative project with US colleagues.

The extra money, which comes from Canada's National Research Council (NRC), will allow the implementation of a bilateral arrangement giving Canadian researchers access to US national radioastronomy facilities on the same basis as US researchers, including the Expanded Very Large Array in New Mexico.

As part of the deal, the NRC will provide some equipment for the array. Through an associated agreement, Canada will also participate in the Atacama Large Millimeter Array to be built in Chile (see *Nature* **422**, 4; 2003).

Homestake strikes gold in underground review

Washington Homestake gold-mine in Lead, South Dakota, has been endorsed by the US National Science Foundation (NSF) as the best site for underground astrophysics and geology experiments.

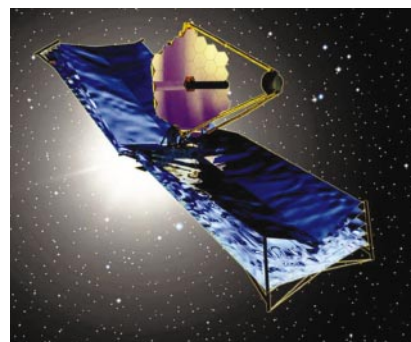
Homestake has been a potential site for an underground lab since it closed as a mine in October 2001. Candidates in last autumn's South Dakota congressional elections promised to win support for the mine, but the NSF insisted on choosing a site by peer review.

Joseph Deheimer, head of the NSF's physics directorate, said last week that the review has now concluded that Homestake is "by far the most favourable site". But the review does not say whether a deep underground laboratory will actually be funded.

Cost-cutting reflected in shrinking mirror

Washington NASA has decided to shrink the primary mirror on the US\$1.6-billion James Webb Space Telescope, the successor to Hubble. The aim is to rein in costs to avoid scrapping one of three onboard scientific instruments (see *Nature* **422**, 3; 2003).

The hexagonal mirror will be reduced to 25 square metres, rather than the nearly 30 square metres proposed by telescope builder TRW, now known as Northrop Grumman Space Technology. Other changes include transferring some responsibilities for building an infrared detector from US to European researchers, and rearranging the funding schedule to provide more money up front.



Downsized: the James Webb Space Telescope will have a smaller mirror than planned.

The European Space Agency, which also has a stake in the telescope, has agreed to provide an Ariane 5 launch rocket. Lift-off is now scheduled for August 2011.

♦ www.ngst.nasa.gov

Game on for PlayStations as consoles take quantum leap

Washington Most teenagers look at video-game consoles and think of gut-splattering shoot-ups. But a team of researchers at the National Center for Supercomputing Applications in Champaign, Illinois, had something else in mind for Sony's PlayStation 2 machine — quantum mechanics.

As part of a project to find cheap ways of doing supercomputing, the team used graphics chips from 70 PlayStations to create a supercomputing cluster. The researchers are now using the chips to tackle quantum-mechanics calculations, and say that other cheap hardware such as graphics cards could also be used in supercomputing applications.

Although the PlayStation network isn't among the world's fastest supercomputers, it is one of the cheapest, at US\$50,000.

Court studies biology papers in row over title

San Diego The fate of key documents from the history of molecular biology, including papers by Rosalind Franklin, is in the hands of the US courts.

Jeremy Norman, a collector who paid around \$1.5 million to build up an archive of molecular-biology documents, filed the lawsuit in a Californian court last month. Norman is suing Al Seckel, a neuroscientist at the California Institute of Technology who helped him to collect the papers, for blocking efforts to sell the archive at auction in April (see *Nature* **421**, 564; 2003). Seckel and some of the researchers who created the documents say that they only sold the papers to Norman on the condition that the collection he was creating would not be split up. Norman is asking the judge to give him clear title to the documents so that they can be sold as he wishes.



Switched off: the Zheleznogorsk nuclear reactor currently produces weapons-grade plutonium.



MANDEL NGAN/AFP PHOTO

Sweet revenge

Parasites exact a devastating toll on health, particularly in the tropics. Could vaccines based on the sugars on parasite surfaces provide a way to fight back? Carina Dennis investigates.



Researchers are studying cultures of *Leishmania* (right) in the hope that its sugary secrets will give respite to the millions suffering from this parasite (above).

They are the undercover agents in the army of pathogens that continually assaults our bodies — parasites that live inside us for months or even years, constantly dodging a gradually tiring immune system, and even burrowing inside our cells.

Various species of *Leishmania*, for instance, can even hide inside some of the immune cells that are supposed to destroy foreign invaders — surviving unscathed in a sac of digesting enzymes that is the cellular equivalent of an acid bath. *Leishmania* causes disfiguring skin sores, and can also result in anaemia, fever and swelling of the liver and spleen. Some 12 million people, mostly in the tropics and subtropics, are currently affected.

Faced with such formidable opponents, vaccine developers have so far drawn a blank — there is currently no vaccine on the market that provides full protection against single-celled protozoan parasites such as *Leishmania*. “Vaccine candidates that work well in animal models have been disappointing in human trials,” observes David Sacks, head of intracellular parasite biology at the US National Institute of Allergy and Infectious Diseases in Bethesda, Maryland.

Part of the problem is that many parasites are able to shuffle their surface proteins rapidly, thereby escaping recognition by the immune system. *Trypanosoma brucei*, a protozoan that causes African sleeping sickness, is the ultimate master of disguise, living in

the bloodstream and replacing its coat of proteins every two weeks.

Given such constantly shifting targets, it is perhaps unsurprising that conventional approaches to immunization — involving whole, killed organisms or purified surface proteins — have yielded little success. So some researchers are now plotting an alternative line of attack: they are trying to get the immune system to respond not to proteins, but to the complex sugars that parasites carry on their surfaces. “Carbohydrates have untapped potential,” says Mitch Kronenberg, an immunologist at the La Jolla Institute for Allergy and Immunology, part of the University of California, San Diego.

Sugar rush

This new focus on complex sugars as candidate vaccines is part of a wider upsurge of interest in glycobiology — the study of the roles played by such molecules in biological systems. Most of the sugars found on cells’ surfaces are linked to proteins or lipids, to form glycoproteins or glycolipids, respectively (see diagram, opposite). Over the past few years, these complex molecules have been found to play crucial roles in important biological processes, including cell communication and signalling.

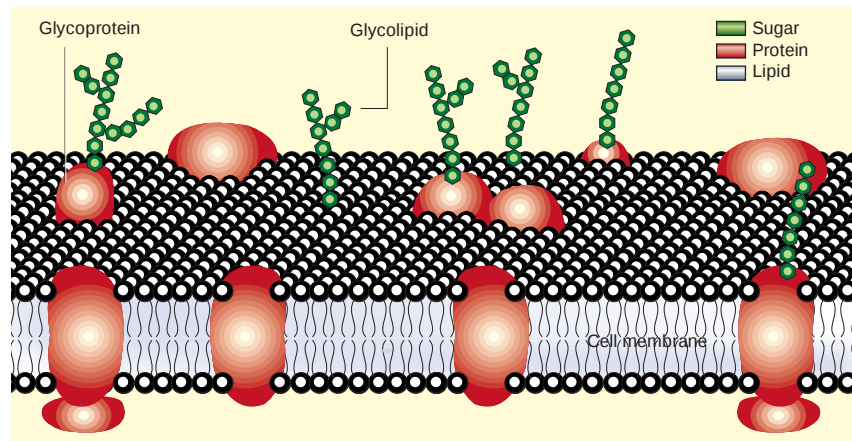
Although the research is in its infancy, there are good reasons to be cautiously optimistic about a sugar-based approach to

vaccine development. Parasites carry sugars that are distinct from those of their hosts — a basic requirement for any type of candidate vaccine molecule. And carbohydrates have an advantage over proteins in that they are less changeable. Many complex sugars are elaborately branched molecules, the construction of which depends on labyrinthine biochemical pathways involving many different enzymes. Whereas a new protein can be produced without major genetic upheaval, that’s not usually true for a complex sugar. “Carbohydrates are more evolutionarily stable,” says Kronenberg. “Parasites can’t change them very easily.”

What’s more, targeting sugars is appealing because they seem to be central to many parasites’ ability to conquer the host’s defences. “Parasites are very dependent on carbohydrates for their survival and infectivity,” says Michael Ferguson, a molecular parasitologist at the University of Dundee, UK. The sugary coats of many parasites, he points out, protect them from harsh environments, help them to invade host cells, and allow them to evade surveillance by the immune system. For example, the surface sugars carried by *Leishmania* are thought to help to protect it from being disintegrated by the enzymes of the intracellular sac in which it resides.

The principle of sugar-based vaccines has already been demonstrated, albeit with less-evasive bacterial infections. Vaccines based on

A. CRUMP/TDR, WHO/SPL



Many proteins and lipids on the surfaces of cells are tagged with sugars to form glycoproteins and glycolipids. Some of these complexes have important functions in cell communication and signalling.

sugars from the surface coats of *Streptococcus pneumoniae*, which causes pneumonia in young children, and *Haemophilus influenzae* type b, which causes a form of meningitis that can result in deafness, brain damage or death, are already widely used. And some early studies provided encouraging signs that a similar approach might work for protozoan parasites — for instance, a glycolipid called lipophosphoglycan (LPG), found on the surface of *Leishmania*, was shown in the mid-1980s to protect mice from subsequent infection¹.

But researchers striving to develop vaccines against parasites have only recently begun to pay serious attention to sugars. The explanation is partly cultural: “Carbohydrates just aren’t as flashy as proteins,” says Sam Turco, a glycobiologist at the University of Kentucky in Lexington.

Technical challenges have also caused sugar vaccines to lag behind their protein counterparts. The biggest problem is that our immune system is not very good at recognizing sugars — and is even worse at remembering them. In contrast, it never forgets a foreign protein, and mounts a more vigorous response the next time that protein is encountered. For this reason, proteins have always been preferred as vaccine candidates.

Our immunological memory depends on the immune system’s T cells, which tend to ignore sugars. But this hurdle can be overcome simply by linking the sugar to a protein. “People tend to shy away from carbohydrates as vaccines because of the widespread dogma that they aren’t very immunogenic,” says Richard Cummings, a glycobiologist at the University of Oklahoma Health Sciences Center in Oklahoma City. “But a carbohydrate can be made very immunogenic with a protein attached.” In the case of the successful bacterial vaccines, the sugar is linked to an inactivated version of the protein toxins produced by the bacteria that cause tetanus or diphtheria.

Scientists have also been deterred from working on carbohydrates because of the molecules’ complexity, which makes them

difficult to synthesize in the lab and difficult to analyze structurally. Whereas the subunits of DNA and protein are strung together in a linear way, like beads on a necklace, carbohydrate subunits can attach to one another at many different points, much as a child’s building blocks can fit together to form a vast assortment of configurations.

High volume

It is also notoriously difficult to extract sufficient quantities of a parasite’s carbohydrate to study. Until relatively recently, researchers had to grow huge vats of parasite cultures just to be able to extract enough of the sugar they were interested in. “That’s why you had graduate students,” quips Turco, who vividly recalls preparing tens of litres of *Leishmania* culture to squeeze out just a few milligrams of a particular glycoprotein or glycolipid.

Now, however, advances in the sensitivity of analytical techniques such as mass spectrometry and nuclear magnetic resonance have allowed sugars to be characterized from much smaller amounts of sample material. “Previously, we’d start with the easiest things first, which meant solving the structures of the most abundant carbohydrates,” says Ferguson. “But as the instrumentation improves, we can tackle the structures of minor cell-surface carbohydrates, which are exquisitely important.”

In parallel, improved methods for making sugars have also made it easier for researchers to study and evaluate vaccine candidates. After determining a sugar’s structure, they no longer have to obtain further samples by purifying the sugar from the parasite itself — which may be difficult to grow in the laboratory. Rather, the sugars can be synthesized artificially. And at the Massachusetts Institute of Technology in Cambridge, a team led by organic chemist Peter Seeberger has streamlined the process

by modifying an automated peptide synthesizer. “The hardware is the same, just the chemistry has changed,” he says. “Carbohydrates that once took years to construct can now be created in a matter of days.”

The device that Seeberger modified is normally used to create specific peptides — the subunits from which complex proteins are constructed — by adding amino acids in succession to form a chain, tethered to a polystyrene bead. Using their adapted device, Seeberger and his colleagues can create complex branched sugars from smaller sugar molecules, determining where each new molecule is added by masking other sites to which it might link with ‘protecting’ chemical groups².

Louis Schofield, a parasite immunologist at the Walter and Eliza Hall Institute of Medical Research in Melbourne, Australia, is now working with Seeberger to develop a sugar-based vaccine to curb the impact of the world’s deadliest parasite, *Plasmodium falciparum*, which is responsible for more than a million deaths every year from

This former protein synthesizer has been modified to build bespoke sugars.





Michael Hewitt (left) and Peter Seeberger have synthesized sugars from a pair of menacing parasites.

malaria. The vaccine won't prevent infection, but Schofield hopes that it will stop the disease progressing to debilitating bouts of fever and to brain swelling — known as cerebral malaria — which can ultimately kill its victims.

The vaccine is based on glycoinositol phospholipid (GPI), a glycolipid produced by *P. falciparum* that is thought to trigger cerebral malaria's characteristic inflammation. Last August, Schofield and Seeberger showed that mice immunized with chemically synthesized *P. falciparum* GPI were much less likely than those in a control group to die or show signs of disease when infected with the related parasite *P. berghei*, which causes cerebral malaria in rodents³. This candidate vaccine will soon be tested in monkeys.

Seeberger and his colleague Michael Hewitt have also synthesized a molecule consisting of four simple sugars that forms the 'cap' of the LPG carried by *Leishmania*⁴. Preliminary unpublished results suggest that it confers immunity in mice, but further development is "on the back burner," says Seeberger, as he and Schofield are currently devoting their efforts to malaria.

The ability to make pure parasite sugars in the lab will also allow researchers to address one of the lingering doubts surrounding early work such as the studies¹ in which mice were immunized with *Leishmania* LPG. "The concern with carbohydrates isolated from biological material was that residual protein contaminants could be stimulating an immune response," explains Ferguson. To address this issue, Ferguson and his Dundee colleague Andrei Nikolaev are now testing chemically synthesized LPG, coupled to proteins or lipids, in immunization experiments in mice.

As well as being deployed against single-celled protozoa, carbohydrate vaccines are also being studied for their potential to offer protection against multicellular parasites. Cummings, for instance, is developing a vaccine for schistosomiasis — a disease caused by flatworms that reside in blood vessels and cause chronic damage to the liver, intestine and bladder. The vaccine is based on sugar chains, called glycans, that are found on

the worms' abundant surface glycoproteins. "Glycans are an ideal target, as they occur in high numbers and on many different glycoproteins," says Cummings. He has already found some evidence that the glycans linked to proteins such as the tetanus toxoid can provoke a strong immune response in mice.

Hidden details

Other parasites pose major problems, however, because their surface sugars are not readily accessible to the immune system. For example, the protein coat of *T. brucei* is so thick that most experts hold out little hope of developing a sleeping-sickness vaccine based on the surface glycoproteins and glycolipids buried beneath. Even here, however, there may be an Achilles' heel to exploit. *T. brucei* propels itself around using a whip-like flagellum, and at the base of this structure is a pocket filled with a gel-like substance through which the parasite absorbs nutrients. "Some



Michael Ferguson says that sugars are crucial for many parasites in beating their hosts' defences.

antibodies can gain access to this pocket," says Ferguson, who is now trying to work out the structures of the carbohydrates in the gel.

With the search for vaccines only just getting under way, some experts think that the best approach will be a wide-ranging fishing expedition for suitable candidates. "We don't really understand parasites well enough to be able to say what is the best molecule to use as a vaccine," argues Rick Tarleton, a parasite immunologist at the University of Georgia in Athens. "One way to proceed is not to make any judgement of what target would be good, and to test everything."

Others, however, argue that it may be more fruitful to narrow the search by gaining a deeper understanding of how the immune system mounts a response to parasite glycoproteins and glycolipids. "Trying to develop vaccines with such limited immunological information is like stabbing in the dark, but maybe some will get lucky," says Turco.

Developing sugar-based vaccines may also require much more information on the function of parasites' complex carbohydrates. "Having the structure of a sugar is one thing, but you need to know what it does," says Carolyn Bertozzi, a chemist at the University of California, Berkeley. An ideal vaccine candidate, she argues, would be a sugar that the parasite needs to infect or disable its host.

In this regard, glycobiologists are at a distinct disadvantage relative to geneticists and protein chemists, who can refer to databases packed with related molecules from which they can gain hints about the function of newly discovered genetic sequences or peptides. When studying parasites, this problem is magnified. "The thing about parasites that slows down the field considerably, compared with studies in other organisms, is that their carbohydrates are so unique," says Turco. "You don't know what to expect."

Some scientists are rising to the challenge, however. Ferguson, for instance, is now trying systematically to knock out key enzymes involved in sugar synthesis in *T. brucei*, to identify the complex carbohydrates that are crucial to parasitic success.

As well as yielding candidate vaccines, projects such as this may also provide a host of targets against which to develop new generations of anti-parasite drugs. Again, it is too early to rate the chances of success, but if the new focus on parasite glycobiology can yield both effective drugs and vaccines, the millions of people whose lives are blighted by these pernicious organisms could look forward to a much sweeter future. ■

Carina Dennis is Nature's Australasian correspondent.

1. McConville, M. J., Bacic, A., Mitchell, G. F. & Handman, E. *Proc. Natl. Acad. Sci. USA* **84**, 8941–8945 (1987).
2. Plante, O. J., Palmacci, E. R. & Seeberger, P. H. *Science* **291**, 1523–1527 (2001).
3. Schofield, L., Hewitt, M. C., Evans, K., Siomos, M.-A. & Seeberger, P. H. *Nature* **418**, 785–789 (2002).
4. Hewitt, M. C. & Seeberger, P. H. *J. Org. Chem.* **66**, 4233–4243 (2001).

Gut reaction

A bacterium that causes ulcers and stomach cancer is on the decline, but not everyone is celebrating. John Whitfield talks to the experts who have misgivings about its impending extinction.



Endangered species: *Helicobacter pylori*, part of our gut flora, may offer clues to human migration.

Save the bacterium! It's an unlikely rallying cry, particularly when the microbe in question causes ulcers and stomach cancer. But *Helicobacter pylori* is in steep decline in many parts of the world, thanks to improved sanitation and the widespread use of antibiotics, and some biologists are beginning to wonder whether its disappearance is really for the best. In the West, the bacterium's demise has been dramatic — half of the US population aged 60 and over are infected with *H. pylori* compared with only 20% of those under 40.

Although most gastroenterologists view *H. pylori*'s disappearance with satisfaction, other researchers point to hints that the bacterium may help to protect against conditions such as infant diarrhoea and oesophageal disease. Some experts say that removing an important member of our intestinal flora will have unforeseen consequences for our inner ecosystem, and so our health. "We have no good sense of the microbial ecology of humans," says infectious-disease specialist Julie Parsonnet of Stanford University in California. "*H. pylori* infection revs up the immune system — what happens to our ability to respond to other infectious agents when that isn't there?" What's more, if *H. pylori* does disappear, so too will an opportunity to use studies of the bacterium to retrace the evolution and migration of human populations.

The debate over *H. pylori* begins with the question of how it got into us in the first place. Most mammals seem to have their own species of *Helicobacter* in their stomachs, leading some researchers to suggest

***Helicobacter pylori* may have been a success because it offers advantages to its host.**

that the relationship between microbe and host predates the evolution of modern humans. But others think that *H. pylori*'s hook-up with humans, and subsequent spread around the world, was more recent.

That's the view taken by geneticist Douglas Berg of Washington University in St Louis, Missouri. Berg and his colleagues looked at more than 500 strains of the bacterium taken from people in five continents. Native Peruvians had bacteria more similar to Spaniards, they found, even though their genes are more similar to those of East Asians — the group that settled the Americas about 12,000 years ago. This led Berg and his colleagues to suggest that the conquistadors may have carried the bacterium to South America about 500 years ago¹, and that the continent's first humans arrived with virgin stomachs.

Martin Blaser disagrees. A microbiologist at New York University, he believes that *Helicobacter* has long been part of the gut flora of all humans. Late last year, Blaser and his colleagues seemed to reaffirm the ancient origins of American *H. pylori*, when they discovered strains closely related to the East Asian version of the bacterium in native people

living in remote regions of Amazonia². And in unpublished work, pathologist Marvin Allison of the Medical College of Virginia in Richmond has detected *H. pylori* in the stomachs of 1,800-year-old Chilean mummies. In areas where Europeans and Amerindians had mixed, Blaser believes, indigenous strains have fallen victim to microbial imperialism.

Berg considers the case to be unproven. Japanese and Chinese adventurers arrived in the New World about the same time as the Spanish, he points out, and they may have been the source of the Asian *H. pylori*. "It's still very controversial," he says. "How one interprets the evidence depends on one's biases."

A family affair

Knowing more about how *H. pylori* spreads would help to solve this puzzle, but so far we have only a rough sketch. Carriers usually pick up the bacterium before the age of ten and stay infected for the rest of their lives. Infected people shed large quantities of *H. pylori* in their diarrhoea and vomit³, which explains the microbe's success in regions with poor hygiene and sanitation. As a result, those at the bottom of the socio-economic scale and people in poor countries are more likely to be infected. The bug most often spreads vertically, from parents to offspring down the generations; there is very little transmission between adults.

Our tendency to keep *H. pylori* in the family means that bacterial and human histories mirror one another. A less faithful microbe that swapped between the members of a generation, a process called horizontal



Imported goods: did the Spanish introduce new bacteria to the stomachs of Amerindians?

transfer, would race around the world in a year or two, detaching itself from the genetics of its host. But *Helicobacter*, which sticks around for decades and — like genes — passes from parent to offspring, could be useful for reconstructing human evolution.

In March, molecular biologist Mark Achtman of the Max Planck Institute for Infection Biology in Berlin and his colleagues published the first global survey of the population genetics of *H. pylori*⁴. Their analysis revealed that the bacterium's genes preserve a record of human migrations both ancient and modern. Recorded through the presence of distinct genetic strains of the bacterium were the arrival of Europeans in South America, waves of migration into Europe from the near East and Central Asia, the settlement of New Zealand by Polynesians, and the influx of Africans into the Americas through the slave trade.

Researchers still debate how faithful this bacterial record of human migrations will turn out to be. Other microbes also seem to carry a record of our ancestors' movements — for example, each continent has its own strain of the human polyomavirus JC virus, another widespread, usually benign infection⁵. Studies of these microbes typically corroborate the evidence from human genetics and anthropology. But should such research suggest alternative hypotheses about our ancestry and movements — if certain strains turn up where they are not expected, for example — it may be difficult to separate misleading interpretations from important new discoveries.

"You have to be careful," says anthropologist Todd Disotell of New York University. As with language and culture, which are also mainly passed down vertically, microbes such as *H. pylori* can sometimes move horizontally, he says. "They're not a perfect record."

Even if *H. pylori* isn't ideal for anthropology, its geographical variance may help us

to understand what makes some people succumb to its harmful effects. In all carriers, *H. pylori* causes a chronic inflammation of the stomach lining; in about 10–15% of carriers this results in ulcers, and about 1–2% will develop gastric cancer, the biggest cancer killer after lung tumours. But these percentages vary around the world. Ulcers and gastric cancer are rare in sub-Saharan Africa, despite a high rate of *H. pylori* infection. In contrast, gastric cancer is most common in the Far East, where genes known to increase the bacterium's virulence are more widespread than in European strains. In India, ulcers most commonly afflict the duodenum; in Japan the stomach is hardest hit.

Tangled web

Differences in bacterial genes do not account for all of these variations. Diet and lifestyle are also important, as is the genetic make-up of the human host⁶. Several groups, including Berg's, are now working to untangle the risk factors, so that efforts to treat the infection with antibiotics can be targeted to those most likely to develop ulcers or cancer.



Poverty trap: poor sanitation in the developing world aids the transmission of *Helicobacter pylori*.

In coming years, treatments other than antibiotics may be necessary, as drug-resistant bacteria are becoming increasingly common. To this end, several research groups are working towards vaccines, some of which are already undergoing clinical trials.

But should we be in such a rush to finish *Helicobacter* off? Blaser argues that *H. pylori* may have been such an evolutionary success because it offers some advantages to its host. For one, it may protect against childhood diarrhoea⁷ — still a major killer across large parts of the developing world — by boosting the immune system and producing peptides that kill other bacteria⁸.

Blaser also notes that *H. pylori*'s wane has corresponded to an increase in acid reflux diseases — serious forms of heartburn — and cancers of the oesophagus⁹. This may be because *H. pylori* can damp down the stomach's production of acid.

But this view is controversial. Most medics view an *H. pylori* infection as an unambiguously bad thing. Some believe that infection with the bacterium may even make childhood diarrhoea worse, as the lowered acid production may allow other bacteria to survive passage through the stomach¹⁰. "*H. pylori* infection is a serious disease," says gastroenterologist David Graham of Baylor College of Medicine in Houston, Texas. "To say it has protective effects is a play on words more than anything else." Graham goes so far as to advocate pre-emptive eradication in regions where gastric cancer is common.

But even without a campaign against it, *H. pylori* seems to be doomed. "In the United States, it's disappearing faster than it would if we had a public-health drive to eradicate it," Parsonnet says.

H. pylori's fate has been easy to track, because it is by far the most dominant bacterium in the stomach. But many of the other 500 or so species of bacteria in our gut might be experiencing population changes without our knowledge. At the moment we can only guess at what the consequences will be. For instance, could shifts in our gut flora have anything to do with the Western world's epidemic of chronic conditions such as allergies and asthma? "Our indigenous organisms are part of our own physiology," says Blaser. "Their extinction may play a role in some of our post-modern diseases."

John Whitfield works in Nature's news syndication team.

- Kersulyte, D. et al. *J. Bacteriol.* **182**, 3210–3218 (2000).
- Ghose, C. et al. *Proc. Natl Acad. Sci. USA* **99**, 15107–15111 (2002).
- Parsonnet, J., Shmueli, H. & Haggerty, T. *J. Am. Med. Assoc.* **282**, 2240–2245 (1999).
- Falush, D. et al. *Science* **299**, 1582–1585 (2003).
- Disotell, T. R. *Genome Biol.* **4**, 213 (2003).
- El-Omar, E. M. et al. *Nature* **404**, 398–402 (2000).
- Rothenbacher, D., Blaser, M. J., Bode, G. & Brenner, H. *J. Infect. Dis.* **182**, 1446–1449 (2000).
- Pütsep, K., Brändén, C.-I., Boman, H. G. & Normark, S. *Nature* **398**, 671–672 (1999).
- Peek, R. M. Jr & Blaser, M. J. *Nature Rev. Cancer* **2**, 28–37 (2002).
- Passaro, D. J. et al. *Pediatrics* **108**, E87 (2001).

SARS, a challenge from the South

When an epidemic threatens the affluent countries, the response is fast and well-funded.

Sir — The next time a major health crisis strikes the world, there must be no doubt as to the motive behind a quick response from the North. This challenge from the South to the world's scientists can be met, for example, by implementing the suggestion of Gerald T. Keusch and Carol A. Medlin in their Commentary (*Nature* **422**, 561–562; 2003) that a global health research network should be established to tackle emerging global crises.

Most scientific breakthroughs are generated in the North and they often neglect the problems that afflict most of the world's people (see, for example, Kofi Annan's "challenge to the world's scientists", *Science* **299**, 1485; 2003). Within a week of receiving samples of the SARS virus in April, Canadian and US scientists independently announced its genome sequence (www.bcgsc.ca/bioinfo/SARS and www.cdc.gov/ncidod/sars/sequence.htm). On 15 May, scientists in Germany announced that available rhinovirus 3CL^{pro} inhibitors might be modified for use in SARS therapy (K. Anand *et al.* *Science* doi:10.1126/science.1085658; 2003).

These results should have an immediate impact on efforts to develop new and rapid diagnostic tests and antiviral agents for SARS. The speed of these praiseworthy efforts was necessary. Yet was their rapid success assured because northern interests were at stake?

Many southern countries are plagued by diseases whose mortality rates exceed that of SARS, for example Ebola virus and sleeping sickness, and the North's reaction to these has been comparatively retarded. This policy is shortsighted and ethically flawed. Although the relatively small size of the SARS genome may have facilitated its rapid sequencing, the recent announcements demonstrate how swiftly the North can react to a major epidemic.

Also in May, US President George W. Bush announced plans to spend more than \$5.6 billion over ten years to build a medical arsenal against biological terrorism. The magnitude of the sum and the relative tempo of the commitment are further indications that, if northern interests are threatened, massive funding can be realized virtually overnight.

There is a stark discrepancy between the standards of health in the developed and developing worlds. More than one billion people in the developing world have yet to benefit from the health improvements of the previous century that are taken for granted in the developed world. This is not to say that the North

has completely ignored southern diseases (see *Nature* **421**, 461–462; 2003). The sequencing of the malaria parasite genome (M. Gardner *et al.* *Nature* **419**, 498–511; 2002) is an example of recognition by the North, in this case the US National Institutes of Health, that eradication of global health inequities is indeed a goal of genomics research.

However, the North must go further, by relaxing or cancelling crippling debt repayments and eliminating subsidies to northern farmers in order to aid southern exporters. Both measures will strengthen southern economies and, it is hoped, result in improved health and social conditions. The challenge for the North is to achieve these goals with the speed and commitment it demonstrated when it tackled SARS.

There are other, pragmatic reasons why the North should undertake these responsibilities. First, southern markets represent untapped growth potential for northern economies that will be frustrated by disease and poverty. Second, southern countries

offer a cheaper labour pool for the North. But illness stifles productivity. Third, neglected southern countries represent a dormant and latent security threat to the North in that illness and poverty beget desperation, instability and crime.

The emergence of SARS highlights the truth that diseases are not necessarily geographically exclusive. Global travel can facilitate the spread of deadly diseases from the South to the North within hours. It is also a warning that continued global health and wealth inequities can ferment new and deadly diseases.

At the same time, the dedication of northern resources to southern diseases must be met with genuine transparency by the South. Corruption, national pride and ego must take a back seat. Only international cooperation can prevent SARS and other emerging diseases from threatening us all. Mother Nature may be the ultimate bioterrorist.

Jerome. A. Singh

Howard College School of Law, University of Natal, King George V Avenue, Durban 4041, South Africa.

Impact factors: target the funding bodies

Sir — Peter A. Lawrence's suggestion (*Nature* **422**, 259–261) for removing the politics from publication is that the well-established scientists should get together and stop over-valuing papers in popular journals. It will be a chilly day in hell before all the older scientists agree to change anything, much less something that currently benefits many of them. A different approach would be more realistic.

The urge to publish in well-known journals is often driven by money. Senior scientists seek top papers to ease the passage of their grants. Search committees are worried about making a huge investment in young scientists, so they seek top papers as a guarantee of future fundability. Tenure and promotions are strongly influenced by the funding that a candidate has managed to raise. The bodies that give grants usually love impact factors as a supposedly quantitative measure of the quality of science. This has inevitably led to overvaluation of the top journals themselves, rather than the science they contain.

We should therefore target the organizations that give grants, rather than hoping to persuade senior scientists to change their behaviour. If funding agencies set out to reward good science, as opposed to

politically successful publications, then the distorted importance of the top journals would be lessened. Of course, productive scientists must be rewarded, so different criteria for achievement would be needed.

My action plan would be to persuade funding bodies to stop using journal impacts entirely, once applicants have been working independently for a few years. Instead, they should ask their referees to answer two questions. First, have this person's published papers been influential to the field? Have they provided the foundations for further discovery, or changed people's perceptions? Or were they predictable or (worse) incorrect? Second, is the applicant's newly published work likely to be influential in this way? Grant bodies have an advantage over journal editors in that they can judge papers retrospectively.

Answers of this sort would not remove the problem completely, as a high-profile paper would always carry better visibility than one in a little-read journal. However, explicit instructions to avoid impact factors — and instead concentrate on the value of past research — could only help to redress the balance.

Robert Insall

School of Biosciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

Editor's note: See last week's issue for other Correspondence on this topic.

Eau de vie

How water shaped our planet and serves as the essence of life on Earth.

Water from Heaven: The Story of Water from the Big Bang to the Rise of Civilization, and Beyond

by Robert Kandel

Columbia University Press: 2003. 312 pp.

\$27.95, £20

Daniel Hillel

Life on Earth began in an aquatic medium, and water is still the principal constituent of all living organisms. In the words of Russian scientist Vladimir Vernadsky, written more than a century ago: "Life is animated water."

Robert Kandel's new book is an attempt to lay out the entire epic story of water from the beginning of the Universe to the present, and even on to the future, all in the space of 300 or so dense pages. As such, it is an audacious *tour de force* that sweeps across time and space at a breathtaking pace. Given the enormous scale and ambition of the undertaking, it is perhaps not surprising that the quality of the coverage is somewhat uneven.

The first two parts — "Water in the Universe from the Big Bang to the Appearance of Man" and "Water in Today's World" — are the most captivating, and demonstrate the author's mastery of such diverse fields as cosmology, astrophysics, geophysics, geochemistry, hydrology, oceanography and climatology. Complex scientific concepts and interrelationships are presented systematically in language that is understandable to the educated lay reader. Kandel displays impressive didactic powers in making intricate concepts clear without resorting to jargon or mathematical equations.

Particularly informative are the explanations of planetary motion (including such phenomena as Earth's spin and obliquity), the Milankovitch cycle, the recurrent ice ages, cloud formations and precipitation, the inter-tropical convergence zone, the El Niño and La Niña oscillations, as well as the natural and the enhanced greenhouse effect. Kandel also sheds light on the interplay between the inherent regularity of the Earth's processes and the mechanisms that cause deviations from that regularity.

Reading and pondering all that leaves one marvelling at the extremely improbable and fortuitous circumstances (including the exact distance from the Sun and the precise values of gravity and ranges of pressure and temperature) that allow water on the surface of our planet to exist in all three interchangeable phases and to make life possible. The discourse is reinforced with a few (one would have wished for more) schematic illustrations.

Among Kandel's insights is an explanation



A splash of colour: jets of hot water spurt out from geysers in Nevada's Black Rock Desert.

of how photosynthesis by aquatic plants enriched the atmosphere with oxygen, which led eventually to the formation of the stratospheric ozone layer. This in turn provided the shield against ultraviolet radiation that allowed life to move onto dry land. He also includes notes about some of the pioneers of the Earth sciences and how the science evolved.

The narrative deteriorates somewhat in the third part of the book, which deviates from strict science into controversial issues relating to the role of water in human history and contemporary affairs. Kandel is right to point out that in many countries (particularly in the arid zone), "agriculture is the biggest water user", and that saving water depends on the "more efficient use of water in agriculture". But he makes no distinction between the processes of evaporation from the soil (which is a loss) and transpiration by plants (a climatically imposed necessity linked to plant productivity), and the concurrent processes of percolation and runoff. The blanket statement that "most of the water used by agriculture returns to the atmosphere by evapotranspiration" is not necessarily true in irrigated agriculture. Where irrigation is inefficient, much of the applied water runs off or percolates through the root zone, raising the water table and causing waterlogging and salination. Only in highly efficient irrigation is nearly all of

the water applied transpired by the crops.

Various other errors are made, too. One is the statement that 500 cubic metres of water per person per year is "an absolute vital minimum", whereas Israel, for example, uses little more than 300 (including water for irrigation) and Jordan even less than that. And it is not only Egypt that "makes intensive use of Nile water"; Sudan uses up to 18.5 billion cubic metres of Nile water per year (one-third of Egypt's usable share of 55.5 billion cubic metres).

More regrettable are the author's digressions into the realm of polemics and his penchant for authoritative expressions of negative opinions regarding "ecologists" (the quotation marks are the author's), who, he thinks, make "completely wrong assertions" that are far too current among environmental movements. He repeats the castigations of "self-styled ecologists" later in the book without specifying exactly to whom and to what they refer. More explicit is his opinion that it is "grotesque" that the US government, "hijacked in 2000–2001 by the oil and coal industries, should maintain that the high-tech US economy depends on wasting crude oil and coal". Such political pronouncements are unbecoming in a science book, which ought to be more objective.

The writing is generally lively and accessible, although the discussion often rambles, with frequent meanderings and *non sequiturs*.

Such minor shortcomings notwithstanding, the book is for the most part a masterful treatment of a complex subject that is of vital and growing importance. All in all, it is a worthy contribution to the popular literature on environmental topics. It teaches much about the primary substance of life and about our utter dependence on it. I recommend the book highly, and the first two sections of it unreservedly. ■

Daniel Hillel is professor emeritus of soil physics and hydrology, University of Massachusetts, and is currently senior visiting fellow at the Center for Climate Systems Research, Columbia University, Goddard Institute for Space Studies, 2880 Broadway, New York, New York 10025, USA.

The last word on books?

A History of Reading

by Steven Roger Fischer

Reaktion: 2003. 384 pp. \$29.95, £19.95

Maurice Pope

Starting from the Bronze Age and ending with modern e-mails and a possible future of e-books, Steven Fischer's *A History of Reading* takes in a wonderful diversity of things. There is a letter to the king of the Hittites from Tutankhamun's widow asking him to let her marry one of his sons; some tenth-century Japanese ladies' literature; reading in the lavatory in twelfth-century Europe; ancient Chinese book-burning; Dr Johnson's advice on what to read and how; Noah Webster's establishment of a patriotically independent American orthography (at a time when for a black man in the South to learn how to read was a crime punishable by flogging or mutilation); statistics on book production; and a discussion of issues such as dyslexia.

But the book is far from being a fact-filled encyclopaedia. It has enthusiasm, opinions (for instance, about how to teach reading and about the problems posed by today's increased accessibility of knowledge), and courage (in declaring that fundamentalism, whether Christian or Muslim, is a threat to civilization, and attacking it accordingly).

It also has direction. "No God revealed everything to us in the beginning," said the Greek poet Xenophanes in the sixth century BC, "but people make discoveries and improvements over time." This was the leitmotiv of Fischer's book on writing, and, perhaps more surprisingly, it also applies to this one on reading. There has been constant progress, not only in the methods of production but in the ease of consumption. Clay, papyrus, parchment, paper and electronic screen represent a ladder of improvement in convenience and cost. So does the change

Philology

The pages of history

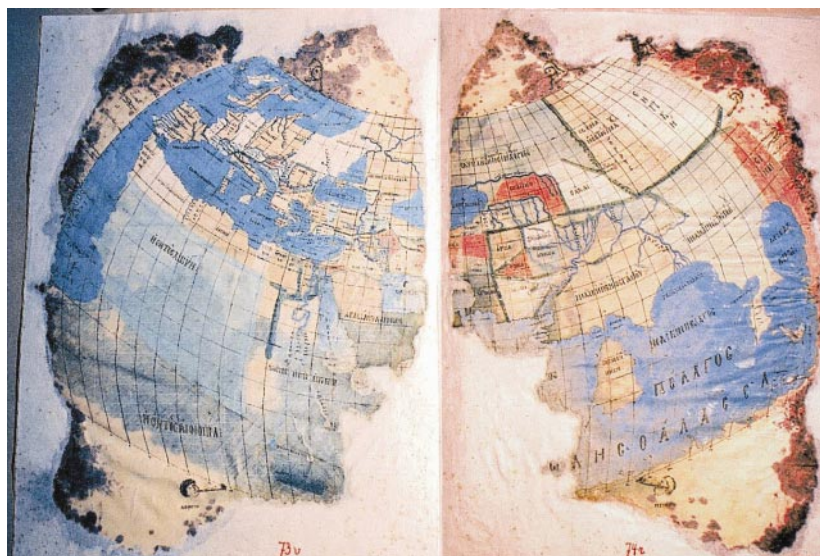
The knowledge of the ancient Greeks found an indirect route to modern European scholars. Works written on parchment or papyrus were serially copied, often unfaithfully, by scribes and translators, as the originals fell apart.

So a copy of Ptolemy's *Geographia* in the original Greek, dating from as early as AD 1300, which has recently come to light in the Topkapi Palace Museum in Istanbul, is causing a stir among philologists. The 120 parchment pages of text and maps, in an admittedly poor state of

repair, are now being examined and annotated by a team at the University of Bern, Switzerland. The team, led by philologist Florian Mittenhuber, is preparing to publish the volume in its entirety, with a German translation.

Differences have already been observed between this Greek text and that of the first Latin translation of *Geographia* from AD 1406, which became the standard in Renaissance Europe and accompanied Christopher Columbus on his voyages to the New World.

Alison Abbott



Old world? This version of Ptolemy's *Geographia* in the original Greek dates from about AD 1300.

from waxed tablet to notepad, and from scroll to book.

There has also been, in our own culture at least, an enormous improvement in presentation. In the Greco-Roman world a scroll (volumen) was much shorter than the average modern book, and a single work might need many volumes. Checking an earlier chapter did not mean just flipping back the pages but a tedious re-winding. There was no table of contents and no index. Worst of all, from our point of view, the text ran solidly without break. There was no punctuation except for a space and a brief line in the margin at the start of a new paragraph. There were not even gaps between words. All of these aids to the reader were introduced, gradually and unmethodically, over a span of some 2,000 years. Furthermore, there was no electric light and there were no reading glasses. For us today, each of these deficiencies would represent a major hurdle, and when we consider them all together we can be amazed at the high literacy rate among our ancestors.

Fischer also sees progress within our own brains. If it is true that all reading was

vocalized until about 1,500 years ago, it may be that the advent of silent reading not only made the process faster but changed its nature. If so, we may be witnessing a new mental faculty in process of evolution.

This theory was first put forward by Julius Jaynes in 1976, and seems somewhat sensational. But adventurous ideas can be forgiven; wrong statements cannot. Fischer says that the Phaistos Disk "apparently conveys a mobilisation proclamation in Minoan Greek". But there is no such language known to scholarship, nor is the decipherment accepted. And the Roman poet Horace did not "boast that his *Ars Poetica* was selling along the Bosphorus, in Spain, Gaul, and Africa" — this would have been extraordinary because the *Ars Poetica* is an essay on literary criticism. What Horace claimed was that his poetry was world-famous. The mistake may be trivial but it could not have been made by anyone who was writing from first-hand knowledge or who verified his references.

Indeed, Fischer relies heavily on secondary sources, more often than not without acknowledging it. If, to take a random sample,

you look at what he says about all of the people in his index whose names begin with A, you will find that on five occasions he acknowledges Alberto Manguel's *A History of Reading* as his intermediate source. On 22 others he gives word for word the same facts and cites word for word the same quotations as Manguel, but without mentioning him. The difference is that in this latter group the original sources are in English or are available in a published translation. Therefore, unless you investigate further, you would think that he had chosen them himself. This seems disingenuous to say the least. Nevertheless Fischer's range is slightly wider than Manguel's, he is more up-to-date by a few years, his material is more systematically organized, and his style is more outwardly objective.

A History of Reading is the third part of an ambitious trilogy by Fischer, the earlier volumes covering the histories of language and of writing. But this volume is self-standing, and can be read and enjoyed for its own sake. ■

Maurice Pope is at 26 Lathbury Road, Oxford OX2 7AU, UK.

New in paperback by the same author **A History of Writing**

by Steven Roger Fischer
Reaktion, £9.95, \$17.95

Darwin's stonecutter

Hewett Cottrell Watson: Victorian Plant Ecologist and Evolutionist
by Frank N. Egerton
Ashgate, 2003, 312 pp. £47.50, \$84.95
Richard Bellon

When the botanist Hewett Cottrell Watson finished reading a presentation copy of Charles Darwin's *On the Origin of Species* (1859), he could not contain his enthusiasm. "You are the greatest Revolutionist in natural history of this century," he wrote excitedly to Darwin, "if not of all centuries."

The book had yet to go on public sale when Watson made this breathless, but prescient, assessment of Darwin's place in the history of science. Darwin looms so large that his shadow eclipses many of his contemporaries, including Watson, who is one of many productive Victorian naturalists who are now largely forgotten. Frank N. Egerton's biography brings Watson out of the shadows and explains why his meticulous work in plant biogeography and systematics was so important to Darwin and other Victorian naturalists.

Recent studies of the history of science have vividly uncovered the political and



Hewett Cottrell Watson was an evolutionist before Darwin but pursued data rather than big ideas.

social background of Victorian evolutionary biology, and yet frequently overlook deceptively mundane technical issues. Nineteenth-century naturalists argued incessantly over such issues as the specific identity of primroses, cowslips and oxlips. Egerton deftly shows how and why such nitty-gritty technical questions mattered in the conceptual development of evolution.

Watson converted to evolutionism in the early 1830s when Darwin was still a non-evolutionist on the *Beagle*. Although Watson's adoption of evolution strengthened his reputation as a renegade, he repeatedly staked out positions that were more radical in principle than in practice. He angrily rejected his father's Tory politics, for example, but not the comfortable gentleman's lifestyle, which was financed entirely by his father's money. He was attracted to phrenology, but his conservative (and sensible) standards of scientific practice later drew him away.

Egerton concludes that Watson, for all his contrarian eccentricity, was no scientific revolutionary. He devoted his energy to the conventional species problems of the age: geographical distribution, the nature and limit of variation, and taxonomic delimitation. His interest in the origin of species remained secondary and unimaginative.

Most botanists before 1859 believed that each plant species had a fixed central type, and that variation from type was a transient response to environmental stimuli. A transplanted species would vary to conform to its new environment, or die if conditions were too alien. Return an environmentally

induced variety to its aboriginal environment, and it would revert to type. Variation, although ubiquitous, could not lead to evolutionary change because it was both inconstant and tightly bounded. Watson accepted this conventional wisdom with a caveat. He suspected that if variation persisted for an unspecified period of time, it became fixed, through some unspecified process, as a new central type. In this way, variation could accumulate to produce new species.

Watson never developed this idea. He believed that the facts of systematics and distribution strongly pointed towards evolution, but also suspected that the origin of species was a question that no one person could solve definitively. He duly subordinated his vague evolutionary speculations to his rigorous systematic and geographical studies. His great treatise on British plant geography, *Cybele Britannica* (1847–59), provided a wealth of suggestive data and analysis but broke no new intellectual ground. Echoing Watson himself, Egerton sees his subject primarily as a stonecutter for Darwin's edifice. By analysing the practical issues driving the development of natural history in the middle decades of the nineteenth century, Egerton allows a deeper appreciation of Darwin's evidence and argument, particularly relating to the chapters in *On the Origin of Species* on "Variation under Nature" and "Geographical Distribution".

The biography does not focus exclusively on Watson's contributions to evolutionary biology. Egerton perceptively analyses Watson's bullying personality, which soured nearly all of his professional relationships. Seeing how Watson's contemporaries balanced his obnoxiousness against his obvious scientific ability provides insight into the moral economy of Victorian science. The chapters on Watson as phrenologist clarify the history of that failed science.

The book disappoints in one key respect. Even though the subtitle characterizes Watson as a plant ecologist, there is little ecology in the book. The term 'ecology' was translated from German late in Watson's life, and he probably never encountered it. Watson's work might have given later ecologists much to use and assimilate, but it still falls squarely into an earlier, and different, tradition.

Egerton opens a window on the complexity and vitality of biology in the age of Darwin. This is a book to pursue, but not to inaugurate, an interest in Victorian natural history, however. Anyone acquainted with the subject, perhaps from Janet Browne's excellent two-volume biography, *Charles Darwin*, will come away with a much clearer appreciation of the vital intellectual milieu that produced Darwin's considerable revolution. ■

Richard Bellon is in the Program in the History of Science and Technology, 148 Tate Laboratory of Physics, University of Minnesota, Minneapolis, Minnesota 55455, USA.

Outdoor pursuits

What was your first experiment as a child?

My mother always took wildflower guides when we went backpacking in the Rockies. We sighed whenever she stopped to identify a flower. Little did I know what I would be doing with the rest of my life...

What single scientific paper or event changed your career path?

As an undergraduate at Pomona College in California, intending to major in literature or history, I took a botany class during which we went into the desert with dissecting microscopes and Philip Munz's *California Flora*. Wow! The complexities of the flora were like a drug — mind-blowing and addictive. I never looked back.

What book has been most influential in your scientific career?

Alfred Russel Wallace's *The Malay Archipelago* and Thomas Belt's *The Naturalist in Nicaragua* both made me want to go to the tropics. I probably would have stayed working on desert plants had it not been for their compelling accounts of diversity and its origins.

What literary character would you employ as a postdoc?

Margery Allingham's Albert Campion. He would find out lots of things, would be mildly amusing and would always do all the driving.

What book currently resides on your bedside table?

Claire Tomalin's *Samuel Pepys: the Unequalled Self*; Marcel Proust's *The Way by Swann's*, the new translation; *Spain's Road to Empire* by Henry Kamen; Joseph Conrad's *The Secret Agent* — and, permanently, *The World of Jeeves* by P. G. Wodehouse.

What music heads the playlist in your car or laboratory?

For peace and tranquillity — Bach's *Goldberg Variations*; The Sixteen singing Anglo-Spanish music from the late sixteenth century (especially Francisco Guerrero's *Ave Maria Sanctissima*); and The Cardinal's Musick recording of masses and motets by Tomás Luis da Victoria. For rollicking jollity, music from the Cape Verde Islands, Louis Mhlanga and The Rolling Stones.

The job of captain on the Starship Enterprise in Star Trek has become vacant. Nominate any real person, living or dead, for the post.

My father, Ed Knapp. He might not want the job, but he certainly has all the qualities I would want if I were a member of the Enterprise crew. He is brilliant at finding his way around: when I was growing up we never

got lost even in the most labyrinthine cities. He listens to everyone's point of view and last, but not least, he is great at fixing things.

Assuming the dead can be raised and/or time travel exists, who from the world outside science would you most like to have dinner with?

Jane Austen — what an observer of human nature. If he would sit down to a dinner conversation with a woman, Samuel Pepys would be a fascinating companion as well.

What one thing would you rescue from your burning laboratory?

A painting of *Magnolia grandiflora* (the southern magnolia) done on vellum by Georg Dionysius Ehret in 1744. When I first saw it, its luminous beauty took my breath away.

What's the best piece of advice you've ever received?

"Publish taxonomic work when it is excellent, for it will never be perfect. Do not do as I have done and wait for that last specimen, just do it." This from Tim Plowman of the Field Museum of Natural History in Chicago who died young, never getting around to publishing his PhD thesis on *Brunfelsia* in the Solanaceae. Ten years after his death, a colleague and I edited and updated it, so it did get published in the end, but Tim never saw it. I like to think he would have approved.

What overlooked or underrated discovery really changed the science in which you work?

The microscope allowed naturalists to look at parts of organisms that were not visible to the naked eye, thus revealing more characters that were of use in classification. The careful examination of characters is what allows taxonomists to pose hypotheses about species identity and relationships — the microscope opened a whole new world of seeing. This is a tool that is as useful today as it was when it was first invented all those centuries ago.

What do you do to relax?

I read novel after novel, often ones I have read before. If I am really stressed and worried, I read mysteries whose endings I already know, or old favourites like Trollope or Dickens. It is very comforting to be among old friends.

What previously under-recognized sport or pastime should be included in the Olympic Games?

Tree-climbing. The great naturalist E. J. H. Corner trained monkeys to climb trees to collect specimens for him, a brilliant solution to the problem of sampling in the rainforest canopy.



Sandy Knapp

Sandy Knapp is a plant taxonomist at the Natural History Museum in London and a specialist on the Solanaceae — the nightshade family. She likes being outdoors (preferably in the tropics), reading and chatting.

What single discovery, invention or innovation would most improve your life?

'Sleeping-Beauty dust' that could be sprinkled liberally on everyone else, so they would all sleep for 100 years and allow me to catch up.

What's the one thing about science that you wish the public understood better?

Just how exciting and dynamic science really is. The dynamism of science often gets lost in the search for final answers. The excitement of being a scientist is partly because you know that any answer you find might get changed or overturned. Science is a journey full of surprises.

What's the most interesting thing in your fridge?

Three frozen grasshoppers in glass tubes and a jar of London snow (collected by my daughter in January).

What would you have become, if not a scientist?

A different scientist — I can't think of anything more exciting or interesting to be.

Name one extravagance you can now get away with because of your eminence.

Eminent people should be allowed to get away with less, not more.

What's just around the corner?

Something that will cause us all to nod and say "I told you so" knowingly — but that none of us really saw coming at all. ■

All change at the synapse

Silvio O. Rizzoli and William J. Betz

Two groups, using different approaches, have provided the first quantitative evidence that neurotransmitter release in the brain often occurs by a long-sought 'kiss-and-run' process.

Synaptic transmission — the transfer of information between nerve cells — is the most important aspect of brain function. Roughly speaking, it works as follows. When electrical impulses propagate to the end of a neuron, they evoke the release of chemicals called neurotransmitters. These diffuse to a nearby neuron and trigger further electrical activity. The neurotransmitters are stored inside the neuron in small, membrane-bounded vesicles¹ — rather like a soap bubble within a larger soap bubble — and they are released en masse through a pore that forms when the vesicle membrane fuses with the neuron's surface membrane. It is well established that the vesicle membrane can then collapse into and coalesce with the surface membrane; after that, it can reform and be reused.

But the retrieval of membrane components from collapsed vesicles is complex², and a different theory, called 'kiss-and-run', envisions no vesicular collapse at all; instead, a fusion pore flickers open and then closes. Although the economy of this concept holds considerable intuitive appeal³, and it is known to occur in secretory processes in other cells⁴, solid quantitative evidence of kiss-and-run fusion at conventional synapses has been virtually absent — until now. On pages 607 and 643 of this issue, Gandhi and Stevens⁵ and Aravanis, Pyle and Tsien⁶ describe how they used different fluorescent markers to track the fate of individual synaptic vesicles, following a single electric shock, in living neurons from the hippocampal region of rat brains. Both groups conclude that a large fraction of neurotransmitter-release events involve kiss-and-run, not vesicle collapse.

Gandhi and Stevens⁵ studied these events by using genetic tools to load synaptic vesicles with synaptophysin, a pH-sensitive fluorescent protein that is attached to a protein in the vesicle membrane⁷. In resting vesicles, the pH is low and the fluorescence of synaptophysin is quenched. But upon exocytosis — vesicle fusion with the surface membrane — the interior of the vesicle becomes open to the solution outside the neuron, allowing protons to escape and so raising the pH. Synaptophysin then fluoresces brightly, and does so until vesicle retrieval (endocytosis) and reacidification occur. Using this approach, Gandhi and Stevens measured the duration of the blips in

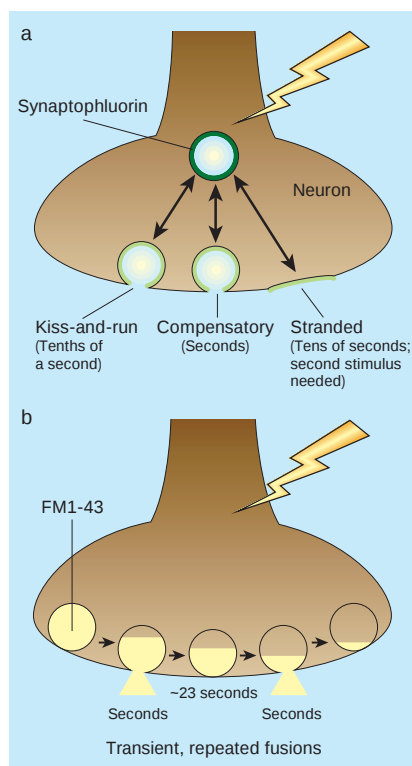


Figure 1 Visualizing vesicles. Upon neuronal stimulation, synaptic vesicles fuse with the surface membrane and release their contents. They are then retrieved. **a**, Gandhi and Stevens⁵ used a fluorescent protein (synaptophysin; green) to study this process; they found that vesicles are retrieved in one of three ways. Some vesicles stay open briefly before retrieval and do not collapse fully into the surface membrane ('kiss-and-run'). Others stay open for longer and probably also do not collapse ('compensatory'). Still others collapse and are not retrieved until another stimulus is delivered ('stranded'). **b**, Aravanis *et al.*⁶ used a dye, FM1-43 (yellow), to study vesicle retrieval. They found that single vesicles can undergo many rounds of fusion. Single electrical stimuli caused only partial loss of dye, which occurred slowly over some 10 seconds. Further release from the same vesicle could sometimes be evoked, but not until a 'dead time' of around 23 seconds had passed. These events probably correspond to the compensatory mode in **a**.

fluorescence that followed single shocks to hippocampal nerves. These blip durations revealed three distinct modes of vesicle retrieval (Fig. 1a).

In one mode, the vesicles remained open for only a few hundred milliseconds. The authors found that this mode involved a kiss-and-run mechanism — that is, the formation of a small fusion pore rather than full vesicle collapse. They discovered this by using two different proton buffers, one that could enter the open vesicle through the narrow neck of the fusion pore, and one that could not.

In the second type of fusion event, the vesicles remained fused with the surface membrane for 8–21 seconds, and were then internalized (the authors call this 'slow', or 'compensatory', endocytosis). Whether the vesicles remained in a kiss-and-run configuration or collapsed completely is unclear from these experiments, but the work of Aravanis *et al.*⁶ (see below) suggests that they did not collapse. Finally, some fusion events persisted for the duration of the observations (45 seconds). A clue to the fate of these

'stranded' vesicles was the finding that, occasionally, the fluorescence actually decreased by one vesicle equivalent in response to a single shock — hinting that endocytosis had occurred in the absence of exocytosis. The frequency of such endocytic events matched the frequency of stranded vesicles, suggesting that stranded vesicles were retrieved in a triggered fashion.

Gandhi and Stevens also analysed the probability of exocytosis occurring at individual nerve terminals in response to a single shock. They found that terminals that responded infrequently operated mostly in the fast kiss-and-run mode, whereas those with a high probability of exocytosis relied mainly on slow, compensatory retrieval.

One limitation of the synaptophysin technique is that a vesicle is visible only while it is connected to the outside. Once it has been retrieved and reacidified it becomes invisible, making it impossible to know, for example, whether two successive exocytic events in the same nerve terminal are produced by the same vesicle or by two different vesicles. Using the approach of

FM-dye labelling⁸, however, Aravanis *et al.*⁶ could continue to track the behaviour of single vesicles. When neurons are bathed in a solution containing the fluorescent dye FM1-43, and then stimulated, any vesicles retrieved by the neurons will take up the dye and become visible under the fluorescence microscope. Further stimulation in a dye-free medium causes dye-loaded vesicles to undergo exocytosis and release their fluorescent contents. With few initial shocks, some synapses acquire only a few labelled vesicles, and individual exocytic events can be observed^{9–11}.

Aravanis *et al.* found that most of the dye-loaded vesicles lost only part of their dye in response to a single shock — as if the fusion pore opened and then closed before all of the dye could escape, as in some sort of kiss-and-run process. Moreover, in 8 of 20 instances of partial dye release, a subsequent shock caused further dye loss (Fig. 1b). This suggests that, in these 8 cases, the vesicle had remained in place after releasing part of its dye load, and had not become detached from the membrane and joined the 30 or so other functional vesicles that have been estimated to exist inside synaptic terminals¹². A 'dead time' lasting some 23 seconds occurred after the first partial release event, when further release could not be triggered. Presumably, vesicles refill with neurotransmitter during this time, and reprime their fusion machinery.

It would be natural to connect these events with those labelled 'kiss-and-run' by Gandhi and Stevens, but this is probably not correct. Gandhi and Stevens' fast kiss-and-run events persisted for only a few hundred milliseconds, too short a time to permit FM dye to escape. By contrast, vesicle collapse during Gandhi and Stevens' 'stranded' events would cause complete loss of FM dye — a relatively rare observation by Aravanis and colleagues. This leaves only the 'compensatory' events. If correct, the results obtained by Aravanis *et al.* (which require a kiss-and-run mechanism) mean that Gandhi and Stevens' slow compensatory events, like the faster events, can be identified as operating by kiss-and-run. Consistent with this, the timing of FM-dye loss⁶ fits well with the time course of compensatory endocytosis⁵.

For decades, the collapse and slow retrieval of synaptic vesicles has been confirmed repeatedly. Yet in the present work^{5,6} a large fraction of events involved a kiss-and-run mechanism, not vesicle collapse. Why the discrepancy? The explanation may lie in the different types of synapse studied. Vesicular collapse has been demonstrated most clearly in preparations such as neuromuscular junctions² and retinal bipolar cells¹¹, which use hundreds of thousands of synaptic vesicles, mobilized from large pools, during repetitive stimulation. Perhaps small synapses in the central nervous system, which operate with far fewer vesicles and do

not need to maintain such high output levels, use a kiss-and-run mechanism to achieve rapid vesicle recycling and reuse, and rely less on mobilizing reserve vesicle supplies.

In future studies it might prove possible to combine the techniques discussed here^{5,6}, using synaptophluorin to determine the time between exocytosis and endocytosis, while simultaneously using FM dyes to probe the size of the fusion pore. If the recent death of Sir Bernard Katz marks the end of a golden age of synaptic physiology¹³, these two studies might be recognized as inaugurating the next chapter in this relentlessly fascinating story. ■

Silvio O. Rizzoli and William J. Betz are in the Department of Physiology and Biophysics, University of Colorado Medical School, 4200 East

Ninth Avenue, Denver, Colorado 80262, USA.

e-mail: bill.betz@uchsc.edu

1. Katz, B. *The Release of Neural Transmitter Substances* (Liverpool Univ. Press., 1969).
2. Heuser, J. E. & Reese, T. S. *J. Cell Biol.* **57**, 315–344 (1973).
3. Ceccarelli, B. & Hurlbut, W. P. *Physiol. Rev.* **60**, 396–441 (1980).
4. Alvarez de Toledo, G., Fernandez-Chacon, R. & Fernandez, J. M. *Nature* **363**, 554–558 (1993).
5. Gandhi, S. P. & Stevens, C. F. *Nature* **423**, 607–613 (2003).
6. Aravanis, A. M., Pyle, J. L. & Tsien, R. W. *Nature* **423**, 643–647 (2003).
7. Miesenböck, G., De Angelis, D. A. & Rothman, J. E. *Nature* **394**, 192–195 (1998).
8. Cochilla, A. J., Angleson, J. K. & Betz, W. J. *Annu. Rev. Neurosci.* **22**, 1–10 (1999).
9. Murthy, V. N. & Stevens, C. F. *Nature* **392**, 497–501 (1998).
10. Zenisek, D., Steyer, J. A. & Almers, W. *Nature* **406**, 849–854 (2000).
11. Zenisek, D., Steyer, J. A., Feldman, M. E. & Almers, W. *Neuron* **35**, 1085–1097 (2002).
12. Harata, N. *et al.* *Trends Neurosci.* **24**, 637–643 (2001).
13. Heuser, J. *Science* **300**, 1248 (2003).

Biogeochemistry

Ancient oceans and oxygen

Matthew T. Hurtgen

The ocean chemistry of 1.5 billion years ago, inferred from rocks of that age, supports the view that marine conditions then were very different from those that pertained at earlier and later times.

Discovering how oxygen levels in Earth's atmosphere and oceans have varied over time is a goal of compelling interest — not least because of the oxygen-dependence of so many life-forms, including ourselves. But how do you estimate an oxygenation state that existed back in deep time? For the oceans at least, sulphate is a crucial indicator, because oxygen largely controls sulphate concentrations. This is because the primary source of seawater sulphate is river input, resulting in part from the oxidative weathering of iron disulphide (pyrite, FeS₂) on land.

Over the past 543 million years, Earth's oceans have largely been well oxygenated — at least well enough to allow the persistence of multicellular life — and rich in sulphate. Before about 2.5 billion years ago, however, they are believed to have been mostly anoxic (without oxygen) and sulphate poor. In 1998, Donald Canfield¹ proposed that for much of the intervening time, the Proterozoic, the oceans had an intermediate oxidation state — that is, the bottom waters were anoxic and sulphidic, a condition known as euxinic, and the surface waters were oxic and contained modest amounts of sulphate. Until now, there has been no concrete evidence for the existence of two chemically distinct water masses during the Proterozoic. But that changes with the appearance of the paper by Shen *et al.* on page 632 of this issue².

The authors present two exciting new data sets from sedimentary rocks in Australia's Northern Territory known as the

Roper Group (Fig. 1). These rocks date to the middle Proterozoic, about 1.5 billion years ago, and their chemistry provides testimony to a continuum of environments that run from onshore to offshore, through near- and far-coastal conditions, to deeper ocean settings. The first set of data, involving iron chemistry, backs up Canfield's ideas. The second set, which uses sulphur isotopes, is consistent with previous hypotheses³ proposing the existence at that time of a sulphate-minimum zone — that is, a region of water along the coast in which sulphate levels were lower than those of the ocean in general — and low sulphate concentrations in the middle Proterozoic ocean.

Just as the weathering of pyrite on land is a major source of sulphate, sulphate reduction and pyrite formation by bacteria in marine sediments — and, under special circumstances, in the water column — are a major sink. Under anoxic conditions, bacteria metabolize organic carbon and reduce sulphate to sulphide, which combines with reactive iron to form pyrite. So the availability of reactive iron can be used to identify the redox conditions of the water column under which sediments were deposited. For example, a measure called the 'degree of pyritization' (DOP) provides an estimate of the extent to which the iron available for pyrite formation was actually transformed to pyrite. High DOP values are typically associated with euxinic settings; low values are characteristic of oxic environments⁴.

Shen *et al.*² have generated DOP data that show systematic differences in iron



Figure 1 Marine exposure: the rocks of the Roper Group in Australia's Northern Territory reveal the oxidation state of Earth's oceans 1.5 billion years ago.

partitioning between relatively shallow, nearshore and presumably deeper, offshore environments in the Roper sediments: nearshore sediments are characterized by low DOP values, offshore sediments by high values. This indication of the existence of chemical gradients between the surface and deep ocean provides compelling evidence that the bottom waters of the middle Proterozoic oceans were euxinic and were overlain by oxic surface waters.

As to the sulphur isotope data, the authors argue that systematic differences in the isotopic composition of pyrite between coastal and oceanic sediments suggest that there were steep lateral gradients in sulphate concentration in the surface ocean. During bacterial sulphate reduction, a significant kinetic isotope effect (given as a value of 4–46‰) occurs as anaerobic bacteria preferentially separate the lighter S isotope — that is, ^{32}S versus ^{34}S — during the production of sulphide and pyrite². For this reason, sedimentary pyrite is commonly isotopically 'depleted' relative to sulphate of the same age.

The extent of this depletion is controlled in part by the availability of sulphate. Under non-limiting sulphate concentrations, sulphate-reducing bacteria have a big pool of sulphate from which to extract the lighter ^{32}S , and larger fractionations (around 45‰) typically occur. By contrast, under sulphate-limiting conditions, two things can happen. First, as sulphate-reducing bacteria preferentially consume the light ^{32}S in forming pyrite, the residual sulphate reservoir becomes enriched in ^{34}S . Second, as sulphate concentrations drop below a certain level, bacteria lose their ability to discriminate between sulphur isotopes, and therefore fractionations decrease⁶. The overall effect of sulphate limitation is a more enriched sulphur isotope signal.

Shen *et al.* discovered that the sulphur isotopic composition of pyrites in the Roper

Basin varies consistently along the palaeo-environmental onshore–offshore transect. Offshore sediments display a wide distribution of sulphur isotopes (around 45‰) that is comparable to maximum fractionation by sulphate-reducing bacteria. The authors attribute this distribution to a combination of pyrite formation in the water column under non-limiting sulphate concentrations, and pyrite formation in the sediments under sulphate-limiting conditions. By contrast, they found that the sulphur isotopic compositions of near-coastal pyrites are more enriched (relative to offshore values) and have a narrower distribution. This implies that bacterial sulphate reduction was occurring in a sulphate-limiting environment and that oceanic sulphate concentrations were diminished in coastal versus offshore surface waters. They suggest

that, for a sulphate concentration gradient to have existed between coastal and oceanic waters, sulphate levels in the middle Proterozoic ocean must have been low.

Overall, Shen and colleagues' iron speciation data are fairly convincing in indicating that the Roper Basin contained euxinic bottom waters overlain by oxic surface waters. As they point out, however, research on other marine sediments of Proterozoic age will be needed to find out how geographically widespread this state of affairs was. And it remains to be seen what produced the systematic differences in sulphur isotopes along the transect. Were the enriched values and decreased fractionations in the nearshore settings indeed the result of a coastal region with a sulphate-minimum zone? Or did they result from pyrite formation in nearshore sediments, where sulphate diffusion was inhibited because the Proterozoic ocean lacked sediment-disturbing macrofauna, and contained uniformly lower sulphate concentrations than today? Whatever the answers, Shen *et al.* have produced a rich collection of data that set the standard for future studies. ■

Matthew T. Hurtgen is in the Department of Geosciences and Penn State Astrobiology Research Center, 801 Deike Building, Pennsylvania State University, University Park, Pennsylvania 16802, USA.

e-mail: mhurtgen@geosc.psu.edu

1. Canfield, D. E. *Nature* **396**, 450–453 (1998).
2. Shen, Y., Knoll, A. H. & Walter, M. R. *Nature* **423**, 632–635 (2003).
3. Logan, G. A., Hayes, J. M., Hieshima, G. B. & Summons, R. E. *Nature* **376**, 53–56 (1995).
4. Raiswell, R. & Canfield, D. E. *Am. J. Sci.* **298**, 219–245 (1998).
5. Kaplan, I. R. & Rittenberg, S. C. *J. Gen. Microbiol.* **34**, 71–91 (1964).
6. Habicht, K. S., Gade, M., Thamdrup, B., Berg, P. & Canfield, D. E. *Science* **298**, 2372–2374 (2002).

Cancer

Out of air is not out of action

Donald P. Bottaro and Lance A. Liotta

Starving cancers of oxygen would seem to be a good way of killing them, but the presence of oxygen-deprived areas in tumours appears to correlate with poor prognosis. A molecular explanation for this has now been found.

When cells are starved of oxygen, they usually die. This is why numerous anti-cancer treatments aim to prevent the growth of blood vessels in tumours, thereby cutting off their oxygen supply. But, in the latest addition to a long list of lessons we are learning about the insidious nature of cancer, it seems that tumours may in fact turn a lack of oxygen to their advantage. Pennacchietti and colleagues¹, writing in *Cancer Cell*, have uncovered a molecular pathway that is switched on by low oxygen levels, and which could cause cancers to become more aggressive

and invade surrounding tissues.

Normal tissues receive a constant supply of oxygen from oxygenated haemoglobin molecules, carried by a continuous flow of blood. When such tissues are subjected to oxygen starvation (hypoxia) — because of a reduction either in blood flow or in the oxygen content of the blood — the eventual result is cell death. Depending on their individual metabolic requirements, some tissues can survive a hypoxic state for longer than others; for example, brain tissue can survive if hypoxia is limited to only a few minutes. To ensure survival, tissues react in two ways:

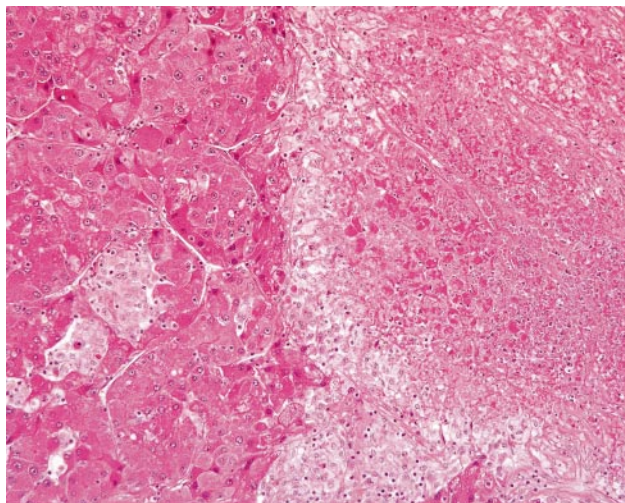


Figure 1 Short of oxygen. The left half of this section of a human tumour had a normal oxygen supply; the right half did not, and is dying. The cells in the right half are shrivelled and fragmented.

they switch into a protective mode by using a specific set of hypoxia-sensing proteins called hypoxia-inducible factors, or HIFs; and they produce 'angiogenesis' proteins that will attract new blood vessels as a way of restoring local blood flow.

Hypoxia has found a place on both sides of the war on cancer. On the one hand, the idea that depriving a primary tumour of essential nutrients and oxygen will stop its growth and spread has won wide appeal. When tumours reach a certain size they outgrow their blood supply, and various studies have linked the resulting hypoxia, via HIFs, to the production of vascular endothelial cell growth factor and consequent stimulation of blood-vessel growth (Figs 1, 2). Following these findings, advocates of anti-angiogenesis strategies have looked for ways to block this molecular cascade by targeting the constituent signalling molecules or the responsive blood-vessel cells (endothelial cells)^{2,3}. On the other hand, several clinical studies⁴ have shown that the presence of hypoxic regions within tumours correlates with poor prognosis and an increased risk of spread to other parts of the body (metastasis), irrespective of treatments used. These findings have cast a different light on tumour starvation as a therapeutic strategy.

Several biological mechanisms have been proposed to explain the observed correlation between hypoxia and accelerated cancer progression. But although there is indirect evidence for these mechanisms, discrete causal links between hypoxia and tumour aggressiveness have been elusive. The report by Pennacchietti *et al.*¹ finally makes a firm molecular connection.

Pennacchietti *et al.* looked at the c-Met protein, which is a receptor for hepatocyte growth factor (HGF), and is found on the cell surface. In a nutshell, they show that when the HIF proteins sense hypoxia, they trigger an increase in the levels of c-Met. The authors first found that hypoxia led to higher levels of c-Met messenger RNA,

indicating protein synthesis, and of c-Met protein in cultured tumour cells. They also identified regions of the c-Met gene that contain typical sequences for binding HIF, and showed that these sequences could be used to drive the expression of a different gene (a so-called reporter gene) in response to hypoxia. Next, the authors turned to grafted tumours in animals and to specimens of human tumours, and found that c-Met and one type of HIF protein were expressed in the same cells. The expression of HIFs is usually taken to indicate hypoxia. In the same samples, the expression of the endothelial-cell marker CD31 and c-Met was spatially distinct. This implies that regions of a tumour that are close to blood vessels (and so probably have normal oxygen levels) express less c-Met.

It is known⁵ that, when HGF binds to

c-Met, it switches on the receptor's intrinsic enzymatic activity and causes it to phosphorylate itself. This addition of phosphate groups stabilizes c-Met's catalytic activity and allows intracellular signalling proteins, such as Gab1, to latch on and also be phosphorylated. Pennacchietti *et al.* now show that, in cultured cells, hypoxia leads to enhanced HGF-induced phosphorylation of c-Met and Gab1. It also results in greater HGF-stimulated cell motility, invasion and branching. But hypoxia-enhanced, HGF-stimulated cell branching is lost when c-Met is repressed.

So, hypoxia leads to increased c-Met levels in tumours, and to increased activity of this signalling pathway (Fig. 2). But why should that influence the aggressiveness of cancer? HGF, the ligand for c-Met, normally stimulates growth, migration and shape changes in a range of cells, including epithelial, endothelial, blood, neural and skin cells, as well as hepatocytes (liver cells)⁶. These diverse effects are crucial to development, organ growth and tissue regeneration^{7–11}. But overactive HGF signalling has also been linked to many different cancers¹¹. For instance, inherited activating mutations in the gene encoding c-Met are associated with renal papillary carcinomas in humans^{12,13}. This is probably partly because activation of the c-Met pathway by HGF stimulates cell division. But it is also known to cause the dissociation of one cell from another, increased cell mobility, and greater production of protease enzymes that degrade the matrix in which cells are embedded — all features that increase the ability of cells to migrate through the extracellular matrix *in vitro*, and which correlate

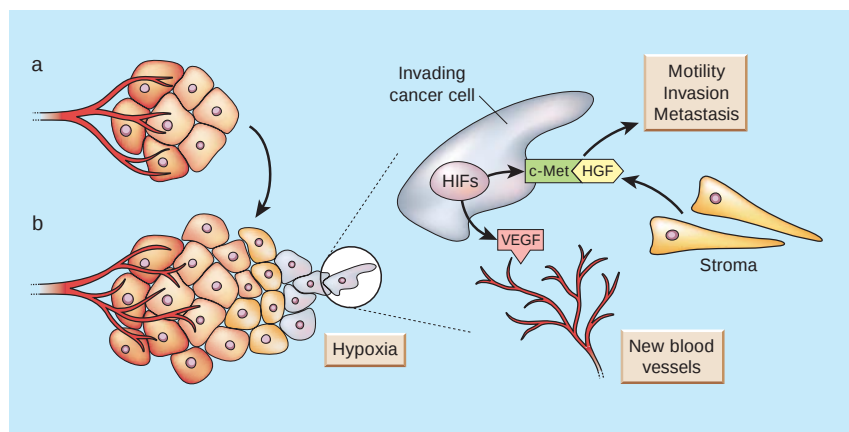


Figure 2 How tumours cope with — and can benefit from — a lack of oxygen. a, At this stage, the growing tumour is well supplied with oxygen and nutrients by its own blood supply. b, The tumour has now outgrown its blood supply and cells at the front line are becoming hypoxic. However, this need not always mean cell death, as is shown on the right. Hypoxia-inducible factors (HIFs) sense the low oxygen levels and stimulate the production of vascular endothelial growth factor (VEGF), a protein that attracts new blood vessels. As shown by Pennacchietti *et al.*¹, HIFs also lead to higher levels of c-Met protein. This protein, on binding hepatocyte growth factor (HGF, which can be produced by nearby stromal cells), can result in increased cell motility, invasion and metastasis. These findings raise concerns about the idea of treating cancer by cutting off its blood supply.

with tumour metastasis *in vivo*¹¹. So the newly discovered connection between hypoxia and c-Met may explain why hypoxia often correlates with a poor prognosis for cancer patients.

This connection also sheds new light on a type of renal-cell carcinoma¹⁴. c-Met is reportedly overexpressed in these carcinomas, which are associated with inherited inactivating mutations in the von Hippel-Lindau (VHL) gene. The VHL protein usually targets HIF proteins for degradation when oxygen levels are normal. So, when VHL is mutated, HIF can no longer be degraded, presumably leading to increased c-Met expression. In effect, the loss of VHL mimics hypoxia, and this can contribute to tumour aggressiveness.

The findings of Pennacchietti and colleagues¹ raise concerns about treatments that aim to cut off a tumour's blood supply. Although the resulting hypoxia might starve the main mass of the tumour, it could also stimulate the motility, invasion and metastasis of tumour cells at the front line — the tumour–host interface. In other words, tumour cells that survive the hypoxia might be selected for increased aggressiveness. But perhaps there is a silver lining to this cloud, if anti-angiogenesis strategies could be

combined with drugs that target the proteins needed for motility and invasion — which now include HGF and c-Met. Such an approach would circumvent the insidious ability of tumour cells to co-opt, and be sustained by, the hypoxia-induced programme of events.

Donald P. Bottaro is in the Urologic Oncology Branch, and Lance A. Liotta is in the Invasion and Metastasis Section, Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892-1500, USA. e-mail: liottal@mail.nih.gov

1. Pennacchietti, S. *et al.* *Cancer Cell* **3**, 347–361 (2003).
2. Ferrara, N. *Semin. Oncol.* **29**, 10–14 (2002).
3. Jain, R. K. *Semin. Oncol.* **29**, 3–9 (2002).
4. Hockel, M. & Vaupel, P. *Semin. Oncol.* **28**, 36–41 (2001).
5. Bardelli, A. & Comoglio, P. M. *Ciba Found. Symp.* **212**, 133–144 (1997).
6. Michalopoulos, G. K. & DeFrances, M. C. *Science* **276**, 60–66 (1997).
7. Birchmeier, C. & Gherardi, E. *Trends Cell Biol.* **8**, 404–410 (1998).
8. Streit, A. C. & Stern, C. D. *Ciba Found. Symp.* **212**, 155–165 (1997).
9. Birchmeier, W. *et al.* *Ciba Found. Symp.* **212**, 230–240 (1997).
10. Funakoshi, H. & Nakamura, T. *Clin. Chim. Acta* **327**, 1–23 (2003).
11. Zhang, Y. W. & Vande Woude, G. F. *J. Cell. Biochem.* **88**, 408–417 (2003).
12. Schmidt, L. *et al.* *Nature Genet.* **16**, 68–73 (1997).
13. Jeffers, M. *et al.* *Proc. Natl Acad. Sci. USA* **94**, 11445–11450 (1997).
14. Maranchie, J. K. *et al.* *Cancer Cell* **1**, 247–255 (2002).

Condensed-matter physics

Hydrogen falls into line

Richard M. Martin and Giulia Galli

Hydrogen ions, both positive and negative, have a pervasive influence on the properties of materials and solutions. A unified picture of the behaviour of these ions has now been drawn.

The role of hydrogen in controlling electrochemical activity is essential to your life and to the life of your computer. The former is well known to any student of chemistry, as the pH of aqueous solutions is defined by the concentration of H⁺ ions. The crossover from a dominance of H⁺ to OH[−] defines the transition from acidic to alkaline solutions, which has profound consequences for chemical and biological processes. Perhaps less well known is that hydrogen is also essential for the operation and lifetime of every modern solid-state electronic device. The ability of hydrogen to take on either of two charge states, H⁺ or H[−], means that it can 'passivate' electrically active defects of either sign in the device material. Introducing hydrogen into the processing makes it possible to achieve the perfection required to reliably produce millions of transistors on a chip. On page 626 of this issue, Van de Walle and Neugebauer¹ propose that these two phenomena are closely related: in fact, they claim to have found a universal alignment of the electro-

chemical level for the change of charge state of hydrogen complexes across a wide range of materials.

Why should there be any relation between the H⁺ and OH[−] ions solvated in liquid water and the hydrogen bound to defects in strongly bonded solids? In the case of water, the role of the solvation shell around the ion has been the subject of studies for many years; the details are still obscure, but interesting suggestions are emerging from simulations that start from first principles^{2,3}. In the case of semiconductors, the mechanism of hydrogen passivation has been elucidated (see, for example, ref. 4), and the structures of the passivated defects vary greatly for different materials. How could such variety be captured in a 'universal' relation?

The proposal made by Van de Walle and Neugebauer¹ is based on calculations done using state-of-the-art methods including 'density functional theory'. The authors are able to predict a large range of formation energies of the individual defect



100 YEARS AGO

The condition of the German Antarctic Expedition which, under the command of Dr. von Drygalski, left Germany in August, 1901, is causing great anxiety... It will be remembered, a correspondent of the *Pall Mall* remarks, that a station was erected on Kerguelen Island in January, 1902, which was intended to serve as a place of observation and as a base for the expedition ship *Gauss*, which was to penetrate much farther south. Those who were at the station, however, suffered terribly from the climate, and then were attacked by beri-beri, which appears to be endemic in that part of the world... The *Gauss* sailed south, but as nothing has been heard of her for a long time it is feared that she is lost, and doubts have been expressed that any of her present officers and crew will ever be heard of again. **ALSO...**

Mr. Marconi is reported to have said on his return to England last week that it will be another six weeks before Transatlantic communication will be resumed. The precise nature of the breakdown has not been published. The American company proposes to extend greatly the system in America by establishing new stations in New York and on the great lakes. It is also stated that the report that Mr. Marconi was suffering from nervous breakdown and would have to take a prolonged rest is unfounded. *From Nature* 4 June 1903.

50 YEARS AGO

When an impulse arrives at a motor end-plate, it causes acetylcholine to be released at numerous 'key points' and in discrete quanta, each containing a large number of molecules. The discharge of each quantal unit is an all-or-none event, and its success or failure is subject to independent statistical fluctuation. Its chances can be altered by various means, for example, by changing the ionic environment. Even under normal conditions, the probability of an individual 'quantum response' appears to be less than unity; that is, a nerve impulse excites at each junction only a fraction of the whole synaptic population. By lowering the calcium and raising the magnesium concentration, the chance of any one unit responding to a nerve impulse can be reduced to a very low value, and under these conditions, which apply to all the experiments described below, the statistical nature of the transmission process can be demonstrated most effectively. **J. del Castillo, B. Katz**
From Nature 6 June 1953.

configurations involving H^+ and H^- ions and neutral hydrogen, H^0 . Under all circumstances, they find that states involving either H^+ or H^- are more stable than those involving H^0 . Whether the H^+ or H^- charge state is more stable depends on the chemical potential of the reservoir of electrons in the material — called the Fermi energy — which determines how easily the conversion between an H^+ and an H^- ion can be made (by adding or removing two electrons). A high Fermi energy means that the system lowers its energy by accepting electrons to become H^- ; as the Fermi energy is lowered, a transition point is reached at which both states have the same energy; beyond this, the H^+ state becomes energetically preferred. Remarkably, Van de Walle and Neugebauer have found that the energy at the transition has almost the same value for a wide variety of materials.

In fact, this universal alignment is not immediately obvious, as there is an additional step to be taken to align the Fermi energy for the transition between H^+ and H^- states in different materials: a reference energy relating electronic states in different materials must be established⁵. In earlier work summarized in ref. 6, Van de Walle and colleagues showed that there is a natural line-up of the band energies at interfaces between materials that can be used to place all the materials on the same scale. Van de Walle and Neugebauer have now calculated the relative positions of the valence-band maximum (the highest energy of the outermost electrons in the atoms of the material) on this scale for a range of materials, including silicon, gallium arsenide and water. Then, when the transition-point energy is plotted on the same scale, the striking constancy of its value from one material to another is clearly seen (Fig. 2 on page 627).

How could such a simple result be found for such varying situations? Indeed, universal rules have been proposed before, for transition-metal impurity states in semiconductors^{7,8}. In some cases there is a natural explanation: the energies obey universal rules because the states are highly localized and hardly change from one host material to another. But for hydrogen the situation is quite different, as its ions form strong bonds and even disrupt molecules or crystalline solids.

A good example of the disruptive influence of hydrogen is seen in crystalline silicon. In its positive-charge state, the H^+ ion prefers to sit in the middle of a Si–Si bond, causing the bond to lengthen. In the negative state, however, the lowest-energy position for an H^- ion is at an interstitial site in an open space in the diamond-like lattice, as far as possible from the Si atoms. In the more typical example of a partially ionic semiconductor, hydrogen breaks an intrinsic bond in the material and forms a new strong bond with

either the anion (in its positive-charge state) or the cation (in its negative-charge state). These two configurations are very different and depend on the semiconductor material. It is even more striking to consider the insulator SiO_2 , in which hydrogen breaks a Si–O bond. Either SiH^+ or OH^- is formed, each complex bonded to other atoms in the network.

Nevertheless, Van de Walle and Neugebauer¹ propose a simple explanation for the universal alignment that they see. In an overall neutral state, the energies of the hydrogen bonds are almost the same, whether hydrogen is bound to the anion or to the cation in a material. The primary factor determining the transition-point energy is the ‘dangling bond’ that remains: the H^- charge state binds to the cation, leaving an unfulfilled, dangling bond on the neighbouring anion (and vice versa for H^+). The dangling bond accepts an electron, or, for H^+ bound to an anion, it gives up an electron. This fixes the transition energy at the ‘mid-gap’ position — effectively an average of the energies to add or subtract an electron — and leads to a natural alignment of transition energies that is not sensitive to the details of the band structure of the material.

This remarkable alignment uncovered by Van de Walle and Neugebauer might prove a

useful guide in understanding more about the physics at interfaces between different materials. There are many examples: the interface between semiconductors and aqueous solutions in photocells; the transfer of protons in fuel cells and hydrogen storage materials; and the interface between computational silicon elements and biological sensors. This work, then, bears not only on the interfaces of materials, but on the interfaces of scientific disciplines. ■

Richard M. Martin is in the Department of Physics, University of Illinois, 1110 West Green Street, Urbana, Illinois 61801, USA.

e-mail: rmartin@uiuc.edu

Giulia Galli is at the Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, California 94550, USA.

e-mail: galligyl1@llnl.gov

1. Van de Walle, C. G. & Neugebauer, J. *Nature* **423**, 626–628 (2003).
2. Marx, D., Tuckerman, M. E., Hutter, J. & Parrinello, M. *Nature* **397**, 601–604 (1999).
3. Tuckerman, M. E., Marx, D. & Parrinello, M. *Nature* **417**, 925–929 (2002).
4. Pankove, J. I. & Johnson, N. M. (eds) *Semiconductors and Semimetals Vol. 34, Hydrogen in Semiconductors* (Academic, Boston, 1991).
5. Van de Walle, C. G. & Martin, R. M. *Phys. Rev. B* **34**, 5621–5634 (1986).
6. Franciosi, A. & Van de Walle, C. G. *Surf. Sci. Rep.* **25**, 1–140 (1996).
7. Caldas, M. J., Fazzio, A. & Zunger, A. *Appl. Phys. Lett.* **45**, 671–673 (1984).
8. Tersoff, J. & Harrison, W. A. *Phys. Rev. Lett.* **58**, 2367–2370 (1987).

Signal transduction

A regulator branches out

Alain Israël

The cellular signalling pathway that leads to activation of the NF- κ B protein has been studied for many years, and one might think that there's little left to learn. But it still has some surprises in store.

Signalling molecules usually contribute to more than one type of process in living organisms, but the NF- κ B family of proteins is particularly multi-talented. It is involved in immune and inflammatory responses, in cell survival and proliferation, and in cancer — and these are just a few of its functions. The NF- κ B proteins are part of a molecular cascade that starts with signals outside a cell and culminates, in the cell nucleus, with the binding of NF- κ B dimers to DNA and the activation of gene expression. It had been thought that some components of this pathway work solely to regulate the movement of NF- κ B to the nucleus. But on pages 655 and 659 of this issue, Yamamoto and colleagues¹ and Anest and co-workers² report a surprising new role for one such component³, an enzyme called IKK- α .

Cells must keep their NF- κ B proteins on a tight leash to prevent them from activating gene transcription wantonly. So, in the

absence of specific extracellular signals, the proteins are kept in the cytoplasm by inhibitors that come in two flavours: the I κ B molecules, which act only as NF- κ B inhibitors, and the p105 and p100 proteins, which serve both as inhibitors and as precursors of NF- κ B DNA-binding subunits (Fig. 1). When cells receive appropriate cues, the I κ B kinase (IKK) complex becomes active and labels the inhibitors with phosphate groups. This phosphorylation leads either to the complete degradation of the I κ B inhibitors, or to the partial degradation of the p100 or p105 proteins. Free NF- κ B dimers are thereby generated, which contain a DNA-binding subunit (p50 or p52) and a transcription-activating subunit, such as p65 or RelB. The dimers then move to the nucleus, where they switch on their target genes.

There are three known subunits in the IKK complex — two protein kinases (IKK- α and IKK- β) and a structural/regulatory subunit (NEMO/IKK- γ)³. Yamamoto *et al.*¹

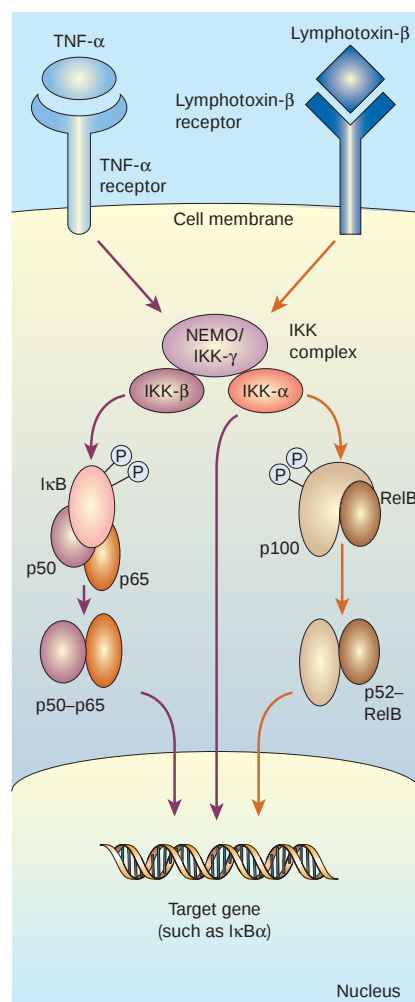


Figure 1 The classical NF-κB cascade, the alternative — and a new twist. NF-κB proteins are dimers, comprising a DNA-binding subunit (such as p50 or p52) and a transcription-activating subunit (such as p65 or RelB). In cells that have not received appropriate external cues, the proteins are kept inactive either by a member of the IκB family in the classical pathway, or by an inactive precursor (in this case, p100) in the alternative pathway. In response to proteins such as tumour-necrosis factor-α (TNF-α) or lymphotoxin-β (top), the IκB kinase (IKK) complex is activated. It phosphorylates IκB and/or p100, leading to degradation of IκB and the processing of p100 into a smaller, p52 form. The p50-p65 and p52-RelB dimers (two forms of NF-κB) then move to the nucleus and activate gene expression. Yamamoto *et al.*¹ and Anest *et al.*² have found that IKK-α can itself move into the nucleus in response to TNF-α, where it associates with certain NF-κB-responsive genes and phosphorylates a histone protein, one of the components of chromatin. How it enters the nucleus and associates with the appropriate genes remains unknown. (Note that the partners of IKK-α in the complex targeted by lymphotoxin-β have not been formally identified.)

and Anest *et al.*² have now identified a wholly unexpected new substrate for the IKK-α component. They show that, in response to extracellular cytokine proteins that influence the inflammatory response — such as tumour-necrosis factor-α (TNF-α) — IKK-α itself moves into the nucleus. Once there, it associates with the promoters (regulatory regions) of several NF-κB-responsive genes, and phosphorylates a component of chromatin.

Chromatin is the compact form of DNA that is found, in association with certain proteins, in the nuclei of eukaryotic organisms; its highly organized structure is important in regulating gene expression. At the heart of chromatin are the histone proteins (H1, H2A, H2B, H3 and H4), which are dynamic components of the gene-transcription machinery. Histones that are located at gene-regulatory regions undergo several types of modification that affect the expression of the corresponding genes⁴. For instance, the phosphorylation of histone H3 at a particular serine amino acid (serine 10) is important in inducing the transcription of so-called immediate early genes (genes that are rapidly turned on and off in response to extracellular signals)⁵. Yamamoto *et al.* and Anest *et al.* now show that IKK-α phosphorylates histone H3 at serine 10.

How does this fit in with what we already know about the NF-κB signalling cascade? Interestingly, it had already been suggested that the two kinases of the IKK complex have different roles. IKK-β is involved in the degradation of the IκB inhibitors in response to pro-inflammatory cytokines such as TNF-α (left-hand pathway in Fig. 1). But IKK-α, besides being involved in an NF-κB- and kinase-independent manner in regulating skin-cell differentiation⁶, is also thought to be part of an alternative pathway. This pathway, activated in response to specific stimuli such as lymphotoxin-β, a member of the TNF family, causes partial degradation of the p100 protein and the movement of NF-κB dimers such as p52-RelB to the nucleus (right-hand pathway in Fig. 1)³. These results have led to the suggestion that IKK-α is not involved in TNF-α-induced activation of NF-κB target genes. But other data have challenged this view⁷, including a report⁸ that it is required for TNF-α-induced expression of the gene that encodes the NF-κB inhibitor IκBα; this gene is also a target of NF-κB.

The new findings^{1,2} suggest that IKK-α is indeed required for TNF-α-induced gene expression — but in a non-classical way. Recruitment to the promoters of NF-κB target genes has so far been formally demonstrated only for members of the NF-κB family such as p65, the major transcription-activating subunit^{9,10}. It has also been reported for more general transcription regulators that associate with NF-κB components¹⁰.

But this is the first time that an upstream component of the signalling cascade has been shown to behave in this way.

There are a few differences between the results obtained by the two groups. Yamamoto *et al.*¹ studied the promoter of the IκBα gene in HeLa cells, and detected IKK-α, but not IKK-β. Anest *et al.*², by contrast, looked at mouse embryo fibroblast cells and found that IKK-β was recruited to the same promoter, with kinetics similar to that of IKK-α. (In fact, they also found that the third component of the complex, NEMO/IKK-γ, was recruited to that promoter, at later times.) But both groups show that IKK-β is not essential for the phosphorylation of histone H3.

These findings raise a number of questions. For instance, what protein (or proteins) conveys IKK-α to the gene promoters? The new data^{1,2} suggest that p65 might be responsible, although the authors did not detect a direct interaction between the two proteins. Also, does IKK-α regulate NF-κB-independent promoters? Does IKK-α-mediated phosphorylation of histone H3 regulate other histone modifications that induce gene expression? What role does IKK-β (and NEMO/IKK-γ) play when it is recruited to chromatin? Do other NF-κB-activating stimuli induce the recruitment of any of these proteins? In this context, Saccani *et al.*⁹ found that IKK-α is not recruited to the promoter of IκBα in cells stimulated with lipopolysaccharide, part of the cell wall of many bacteria — implying that such recruitment is not a general mechanism. In another seemingly contradictory result, Cao *et al.*¹¹ showed that cells with an inactive version of IKK-α respond normally to TNF-α, interleukin-1 and lipopolysaccharide. But the regulation of specific NF-κB-responsive genes such as IκBα has not been studied in this genetic setting.

Whatever the answers to these questions, the discovery that IKK-α has this activity will open up a new avenue of research into the NF-κB cascade, bringing the nuclear events associated with activation of this signalling pathway back into focus.

Alain Israël is at the Unité de Biologie Moléculaire de l'Expression Génique, URA CNRS 2582, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France.

e-mail: aisrael@pasteur.fr

1. Yamamoto, Y., Verma, U. N., Prajapati, S., Kwak, Y.-T. & Gaynor, R. B. *Nature* **423**, 655–659 (2003).
2. Anest, V. *et al.* *Nature* **423**, 659–663 (2003).
3. Ghosh, S. & Karin, M. *Cell* **109**, S81–S96 (2002).
4. Strahl, B. D. & Allis, C. D. *Nature* **403**, 41–45 (2000).
5. Cheung, P., Allis, C. D. & Sassone-Corsi, P. *Cell* **103**, 263–271 (2000).
6. Hu, Y. *et al.* *Nature* **410**, 710–714 (2001).
7. Li, X. *et al.* *J. Biol. Chem.* **277**, 45129–45140 (2002).
8. Li, Q. T. *et al.* *Genes Dev.* **13**, 1322–1328 (1999).
9. Saccani, S., Pantano, S. & Natoli, G. *J. Exp. Med.* **193**, 1351–1359 (2001).
10. Zhong, H., May, M. J., Jimi, E. & Ghosh, S. *Mol. Cell* **9**, 625–636 (2002).
11. Cao, Y. *et al.* *Cell* **107**, 763–775 (2001).

Photonics

Light gets a useful shock

Phys. Rev. Lett. **90**, 203904 (2003)

A shock wave passing through a photonic crystal does strange things to light, say Evan J. Reed and colleagues. In a photonic crystal, a periodic variation in the dielectric constant (caused, for example, by a lattice of holes) creates a 'photonic bandgap' — a range of frequencies within which light is excluded from the material. Light striking a shock-wave front in such a material is reconstituted in ways that have no parallels in other systems.

First, it is shifted in frequency: red light might bounce back as green. Changing the colour of light is important for photonic technologies, and is usually done by passing high-intensity light through a nonlinear optical material. But in the photonic crystal, the shift happens at any intensity, and it can be tuned by altering the frequency of the photonic bandgap.

Furthermore, the light is temporarily trapped at the shock front before being re-radiated. The ability to 'stop' and delay light is useful for optical telecommunications, and the delay here is dependent on the crystal structure and shock speed. Finally, the bandwidth of the reflected light is narrowed without any energy loss, making it more 'monochromatic' — another technological boon if it can be made to work. **Philip Ball**

HIV

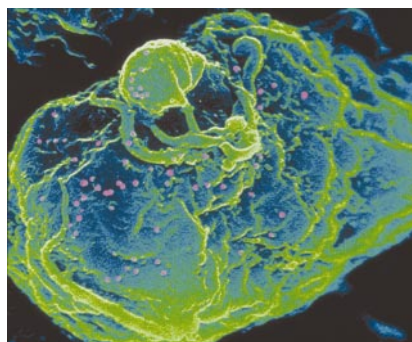
Lucky mutation

J. Clin. Invest. **111**, 1547–1554 (2003)

Some HIV-infected people do not develop full-blown AIDS, despite carrying detectable levels of replicating virus. These lucky few may owe their good fortune to a mutation in the viral protein R (Vpr) gene that helps to protect the immune system's T cells (pictured) from death.

Julian J. Lum and colleagues have found that up to 80% of 'non-progressing' patients are infected with HIV that carries this mutant gene. The authors also genetically engineered strains of HIV to contain either the normal or the mutant form of the gene. Cultured T cells infected with the 'normal' virus died in large numbers. But cells treated with the mutated virus were harder. T-cell numbers also fell less steeply in mice injected with the mutant Vpr protein compared with those injected with the normal protein. And the mutant protein induced fewer of the signs of cell death in mitochondria, the cellular powerhouses.

Patients with rapidly progressing HIV infections show high levels of T-cell death, whereas non-progressing patients have



Under attack: a T cell infected with HIV (small pink dots).

levels similar to those of uninfected people. Lum *et al.* conclude that Vpr is likely to play a significant part in T-cell depletion in HIV-infected individuals — a discovery that may lead to the development of anti-HIV treatments based on Vpr inhibitors.

Helen R. Pilcher

Climate change

Hot tips on temperature

Geophys. Res. Lett. doi:10.1029/2003GL018667 (2003)

The Earth could get warmer faster, according to a new model that takes an 'Earth systems' approach to climate change.

The model, produced by Chris D. Jones and colleagues, factors in a wealth of influences on climate that had not been included in global-change models before — such as an interaction between the Earth's carbon and sulphur cycles, emissions from volcanoes, and the effects of atmospheric ozone and anthropogenic sulphates.

The model assumes that, as the planet warms, carbon will start to be released from terrestrial sinks, such as the Amazon basin. By 2100, this feedback effect should result in 7 billion tonnes of carbon being released into the atmosphere each year and a rise in average global temperatures of 5.5 °C — around 1.5 degrees higher than without the feedback mechanism. A similar increase has been predicted in previous work, but the present analysis is more robust as, working backwards, it accurately models the warming that was observed during the twentieth century.

The possible contribution of carbon-cycle feedback to future warming is hotly disputed, but the use of this 'Earth systems' approach in competing models should help to resolve the question.

Tom Clarke

Plant biology

Buttercup bondage

Am. J. Bot. **90**, 724–729 (2003)

According to Candace Galen and Maureen L. Stanton, the reproductive success of the high-altitude-dwelling snow buttercup depends on its ability to keep the sun in

its sights. Time spent basking in the sun's rays improves the performance of the flower's pollen.

Snow buttercup (*Ranunculus adoneus*) flowers move throughout the day so that they stay closely aligned with the sun's rays. This ability is thought to be useful for luring pollinators: the flower's waxy petals reflect the sun onto the pollen-receiving pistil, creating a warm, spa-like environment that attracts insects. Now, using a bit of buttercup bondage, Galen and Stanton have shown that solar tracking is more than just an insect-tease.

In a series of studies to test the effects of tethering the flowers in place, the authors show that pollen quality (a paternal trait) and pollen receipt (a maternal trait) improve with solar tracking — perhaps reflecting better 'microenvironments' in the parental flowers. Pollen from solar-tracking flowers was more likely to germinate than pollen from tethered flowers. And when pollen recipients were left free to follow the sun, the number of germinating grains per pistil increased.

Hannah Hoag

Quantum chemistry

At full stretch

Angew. Chem. Int. Edn **42**, 2251–2253 (2003)

Can chemistry be done mechanically? Can we make molecular structures simply by pulling atoms into new arrangements? Daniel Krüger and colleagues use quantum-mechanical computer simulations to discover the effect of tugging on a tiny cluster of gold atoms. They find that the mechanical force induces progressive rearrangements from a compact shape to an elongated atomic chain that finally snaps.

These effects have been inferred from earlier experiments in which the tip of a scanning probe microscope is pressed into a gold surface and then withdrawn, pulling out a gold 'nanowire'. And Krüger and colleagues have shown previously that a single thiolate molecule, attached to the probe tip, can act as a hook that forms a single bond to the gold surface. But what if, instead of a flat surface, one pulls on a cluster of just a few gold atoms?

This produces various discrete 'molecular' configurations, each locally stable while having a substantially higher energy than the initial compact cluster. Some of them rearrange suddenly as they are extended, producing a sawtooth force-extension curve. Such structural changes can be regarded as 'mechanoisomerization', comparable to the more familiar thermal or photochemical isomerization.

Philip Ball

The double puzzle of diabetes

Jared Diamond

Why is the prevalence of type 2 diabetes mellitus now exploding in most populations, but not in Europeans? The genetic and evolutionary consequences of geographical differences in food history may provide the answer.

Type 2 diabetes mellitus exacts a huge toll in money and human suffering. For instance, it accounts for more than 100 billion dollars of healthcare costs annually in the United States, or 15% of costs due to all diseases combined. The number of cases worldwide is estimated at 150 million. But this is a minimum number because, for each diagnosed case, there is thought to be one undiagnosed case in First World countries and eight in the Third World¹. Despite its other name of adult-onset diabetes, the disease is becoming more common in young people^{2,3}. At its present rate of increase, within a few decades it will be one of the world's commonest diseases and biggest public-health problems^{2,4}, with an estimated minimum of half-a-billion cases⁵. This explosion in prevalence is occurring especially in the Third World, at about 50% per decade; and because the epidemic is just beginning in the world's two most populous countries, India and China, by the year 2010 more than half of the world's diabetics will be Asians^{2,3,5}.

There are two main forms of diabetes mellitus, and the principal characteristics are outlined in Box 1. But this is the story of type 2, not only because it is much more common and rising steeply in prevalence, but because that rise is doubly puzzling. Not only would the disease seem to be highly disadvantageous in terms of natural selection, but some human populations are much more affected than others.

The geographical variations are shown in Fig. 1 (overleaf). The lowest prevalences, of practically zero, are in rural Third World areas, whereas the highest, 37–50%, are among Nauru Islanders of the tropical Pacific^{6,7}, Pima Indians in Arizona⁸ and urban Wanigela people in Papua New Guinea⁹. Most of the world's broad geographical groupings of people include populations of both very low and very high prevalence — for instance, Mapuche Indians versus Arizona Pima Indians, and rural New Guineans versus the urban Wanigela. Populations undergoing increases in the incidence of type 2 diabetes include not only Asian Indians and Chinese, but also Japanese, Aboriginal Australians, Hispanic Americans and Afro-Americans^{2,4,10}. A conspicuous exception is the absence of any comparable explosion, or very-high-prevalence population, among people of European ancestry. Thus, the puzzling aberrations are not Nauruans and

Pimas, as usually assumed; they are merely the extreme examples. Instead, the aberration demanding explanation is Europe.

As an evolutionary biologist, I have long been puzzled by these differences. In this article I shall suggest a hypothesis for why we are not seeing nearly as much of an explosion among Europeans as among other populations. The evidence comes initially from food history, and tests of it may come from medicine, medical history and molecular biology.

Genetics and lifestyle

Genetics The high prevalence of type 2 diabetes in any large population poses a further evolutionary question. Why is the disease so common, when it should disappear as those genetically susceptible to it are removed by natural selection? (Readers who answer, "Because it kills only older individuals whose child-bearing or child-rearing years are behind them" will find this objection answered below.) The disease certainly has a genetic component^{11,12}, as is evident from the following.

- There is a concordance in diagnosis of nearly 100% for monozygotic twins (those that develop from the same fertilized egg, and so have identical genetic constitutions), but only 20% for dizygotic twins. The latter figure is comparable to that for non-twin siblings, suggesting that factors in the uterus

play a quantitatively minor role (without denying the existence of such factors^{3,13}).

- The prevalence of diabetes among Hispanic Americans varies according to their proportion of Native American genetic ancestry¹⁰.

- Many specific genetic susceptibility factors have been identified^{11,12}. Because the highest mutation rates for any human gene are only around 10^{-5} per generation, the expectation is that only deleterious genes with prevalences below 10^{-5} could be sustained within a population by recurrent mutations alone. The actual incidence of type 2 diabetes is up to 50,000 times higher. And the high prevalence of the disease in many large, ancient, well-mixed populations rules out explanations in terms of the founder effect or genetic drift. These, respectively, are instances where there was only a very limited initial genetic variability, or where random processes such as selective extinction operated, and they tend to affect only small populations. Hence, the high prevalences of type 2 diabetes, like the prevalence of sickle-cell anaemia in certain groups, must be sustained by some compensating advantage that offsets the obvious morbidity and mortality (in the case of sickle-cell disease, the compensating advantage is a certain resistance to malaria).

Lifestyle In addition to that genetic component, type 2 diabetes also involves environmental and lifestyle risk factors —

Box 1 The diversity of diabetes mellitus

The term 'diabetes mellitus' covers a wide variety of conditions that are linked only by shared symptoms arising from high levels of blood sugar. That diversity may be crudely partitioned^{2,3,11,12} into type 2 (adult-onset) and the less-common type 1 (juvenile-onset). The respective prevalences among diabetics in the United States are 90–95% and 5–10%. Both diseases centre on the hormone insulin, which is responsible for mediating the uptake by cells of glucose from the blood.

Type 1 diabetes (insulin-dependent diabetes mellitus) is an autoimmune disease in which autoantibodies destroy the pancreatic-islet cells that synthesize insulin. Patients are thin, produce little or no insulin, and are prone to ketosis, a particular metabolic imbalance. They carry certain gene types — the HLA alleles *DR3*, *DR4* or both — that encode particular components of the immune system. Type 2 diabetes (non-insulin-dependent diabetes mellitus) involves altered insulin secretion and insulin resistance. Patients are often obese and are not subject to ketosis. They do produce insulin but become insulin-resistant — that is, unable to respond effectively to it.

Distinguishing the two forms can be complicated, however, because there is early-onset type 2 and late-onset type 1. Type 2 diabetes is itself very heterogeneous, both genetically and in the associated pathological and physiological symptoms. The disease arises from at least 60 identified genetic disorders, united only by the common feature of high blood-glucose levels due to insulin resistance. This heterogeneity reinforces the evolutionary puzzle: genes that predispose the bearer to type 2 diabetes must really convey some advantage, because they have evidently been preserved independently many times by natural selection. The 'thrifty gene' hypothesis, according to which such genes allow efficient food utilization in times of plenty, in preparation for famine, provides a possible explanation.

J.D.

especially, high calorie intake and low exercise^{2,11,12,14}. For example:

- Disease prevalence is 5–10 times higher in obese people than in those of normal weight.
- The metabolic abnormalities and symptoms of diabetes can often be reversed by dieting and exercise.

- The symptoms decline or disappear in populations under starvation conditions — as, for instance, they did in French diabetics under the food rationing imposed during the 1870–71 siege of Paris³.

- Prevalence of diabetes increases within about two decades^{1,15} in populations that have adopted a high-calorie, low-exercise lifestyle as a result of emigration: for instance, when Yemenite Jews were airlifted to Israel, and when there was a burst of Japanese emigration to the United States. Other examples are groups of emigrant Asian Indians in Fiji, Mauritius, Singapore, Tanzania, the United States and Britain^{3,4,16}, and of emigrant Chinese in Hong Kong, Mauritius, Singapore and Taiwan^{2,3,17}.

- Prevalence similarly rises rapidly in a population that remains in the same geographical area but in which calorie intake increases and exercise decreases. Examples include the Nauru Islanders, Arizona Pima Indians, white Australians, urban Aboriginal Australians, urban black Africans in Cape Town¹, urban Samoans, Chinese, Asian Indians and Japanese.

- In Japan, graphs against time of the incidence of type 2 diabetes and of economic indicators are parallel — down to details of year-to-year wiggles. That's because people eat more, and so risk developing diabetic symptoms, when they have more money³.

- As I learned while serving on the Animal Regulation Committee of the Los Angeles Zoo, there is now a diabetes epidemic among captive populations of many primate species whose zoo lifestyle approximates the high-calorie, low-exercise lifestyle of urban humans in the First World. It will be instructive to see whether or not our primate relatives share genes conferring susceptibility to diabetes with us.

All in all, the basis of type 2 diabetes can be summarized as follows³: it “is a lifestyle disorder with the highest prevalence seen in populations that have a heightened genetic susceptibility; environmental factors associated with lifestyle unmask the disease”.

Nauru

Nauru is a remote island in the Pacific that was colonized by the Micronesians in pre-historic times. It was annexed by Germany in 1888, occupied by Australia in 1914 and eventually achieved independence in 1968. It is the world's smallest republic, but it also has a less welcome distinction. The island is the site of a grimly instructive epidemic of diabetes, which illustrates a rarely documented

phenomenon — an epidemic of a genetic disease^{6,7}. Epidemics of infectious diseases wax when transmission of the infectious agent increases; they then wane when the number of susceptible potential victims falls, due both to acquired immunity of the survivors and to differential mortality of those who are genetically susceptible. An epidemic of a genetic disease waxes because of a rise in environmental risk factors, and then wanes when the number of susceptible potential victims falls (but only because of the preferential deaths of those who are genetically more susceptible).

The traditional lifestyle of Nauruans was based on agriculture and fishing, and involved frequent episodes of starvation because of droughts and the island's poor soil. Early European visitors nevertheless noted that Nauruans were plump, and that they admired big, fat people and put girls on a diet to fatten them and so make them more attractive. In 1906 it was discovered that most of Nauru consists of high-quality phosphate rock that could be used for fertilizer, and in 1922 the mining company extracting the rock began to pay royalties to the islanders. As a result of this new wealth, average sugar consumption by Nauruans reached a pound per day by 1927, and labourers were imported because Nauruans disliked working as miners.

During the Second World War the island was occupied by Japanese military forces, who imposed forced labour, reduced food rations to half-a-pound of pumpkin per day, and then deported most of the population to Truk, where half of them died of starvation. When the survivors returned, they regained their phosphate royalties, and resumed eating sugar and other store-bought food. They abandoned agriculture almost completely, became sedentary, and came to rely on motor vehicles to travel around their 20-km² island. Following independence in 1968, per capita phosphate royalties rose to A\$37,500 (US\$22,500) annually, making Nauruans among the world's richest people. Today they are the most obese and have the highest blood pressure of all peoples in the Pacific; their average body weight is half as much again as that of Australians of European origin.

Although colonial European physicians on Nauru knew how to recognize type 2 diabetes, and diagnosed it there in non-Nauruan labourers, the first case in Nauruans was not noted until 1925. The second was recorded in 1934. After 1954, however, the prevalence of the disease rose steeply and it became the commonest cause of non-accidental death. One-third of all Nauruans over the age of 20, two-thirds of those over age 55, and 70% of those few who survive to the age of 70, are diabetics. Within the past decade, prevalence of the disease has begun to fall, not because of mitigation of environmental risk factors

(obesity and the sedentary lifestyle are as common as ever), but presumably because those who are genetically most susceptible have died. If this interpretation is correct, then Nauru provides the most rapid instance known to me of natural selection in a human population — an occurrence of detectable population-wide selection within less than 40 years.

The case of Nauru also illustrates why, earlier in this article, I dismissed the usual objection that type 2 diabetes lacks selective impact because it supposedly affects people only when their reproductive years are behind them. In fact, although the disease appears mainly after age 50 in Europeans, in Nauruans and other non-Europeans it affects people of reproductive age in their twenties and thirties, especially pregnant women, whose fetuses and newborn babies are also at increased risk. For instance, in Japan today, more children suffer from type 2 than type 1 diabetes, despite the latter's popular name of juvenile-onset diabetes. Moreover, in traditional human societies, unlike modern First World societies, no old person is truly 'post-reproductive' and selectively unimportant, because grandparents contribute crucially to the food supply, social status and survival of their children and grandchildren¹⁸.

Thrifty genes

The leading evolutionary theory for the possible benefits of genes predisposing to type 2 diabetes is James Neel's 'thrifty gene' hypothesis^{3,14,19,20}. Neel postulated the existence of metabolically thrifty genes: these permit more efficient food utilization, fat deposition and rapid weight gain at occasional times of food abundance, thereby making the gene-bearer better able to survive a subsequent famine. Examples of thrifty genes would include those resulting in high levels of insulin or of leptin (a hormone released by fat cells that regulates appetite), or in hair-triggered insulin release. Such genes would be advantageous under the conditions of unpredictably alternating feast and famine that characterized the traditional human lifestyle, but they would lead to obesity and diabetes in the modern world when the same individuals stop exercising, begin foraging for food only in supermarkets, and consume three high-calorie meals day in, day out. Following Arthur Koestler, Zimmet refers to the spread of this lifestyle to the Third World as “coca-colonization”^{3,4}.

So accustomed are we in the First World to regular meals that we find it hard to imagine the fluctuating food availability that was formerly the norm and remains so in some parts of the world. I often encountered such fluctuations during my fieldwork among New Guinea mountaineers still subsisting by farming and hunting. For example,

some years ago, in a memorable incident, I hired a dozen men to carry heavy equipment all day over a steep trail up to a mountain campsite. We arrived just before sunset, expecting to meet another group of porters with food, and instead found that they had not arrived because of a misunderstanding. Faced with hungry, exhausted men and no food, I expected to be lynched. Instead, my carriers just laughed and said, "*Orait, i samting nating, yumi slip nating, enap yumi kaikai tumora*" ("OK, it's no big deal, we'll sleep on empty stomachs tonight and wait till tomorrow to eat"). Conversely, on other occasions when pigs were slaughtered for a feast, the New Guineans would consume prodigious amounts of food. This anecdote illustrates an accommodation to the pendulum of feast and famine that was very necessary in times when that pendulum swung often but irregularly — a situation that was much more typical of our evolutionary history than the state of plenty to which we are accustomed.

Two lines of human evidence and two animal models support the plausibility of Neel's thrifty gene hypothesis. Non-diabetic Nauruans and Arizona Pima Indians have postprandial levels of plasma insulin (in response to an oral glucose load) that are triple those of Europeans¹⁴. And given ample food, diabetes-prone populations of Pacific Islanders, Native Americans and Aboriginal Australians do exhibit more propensity to obesity than Europeans: first they gain weight, then they develop diabetes. As to the animal examples, laboratory rats carrying genes predisposing them to type 2 diabetes and obesity survive starvation better than do normal rats, illustrating the advantage of these genes under occasional conditions of famine¹⁶. And the Israeli sand rat, which is adapted to a desert environment with frequent scarcities of food, develops high levels of leptin and insulin, and insulin resistance, obesity and diabetes, when maintained in the laboratory on a 'westernized rat diet' with abundant food. But those symptoms reverse when its food is restricted²¹.

Natural selection

The thrifty gene hypothesis provides an explanation for why humans in general become prone to diabetes under a westernized lifestyle. But why, in light of this hypothesis, are Nauruans and Arizona Pima Indians experiencing especially severe epidemics of type 2 diabetes while European populations — in Europe and elsewhere — have uniquely low prevalences?

Nauru Nauruans suffered two extreme bouts of natural selection for thrifty genes, followed by an extreme bout of coca-colonization⁷. First, like other Pacific Islanders — but unlike the inhabitants of continental regions — their population was

Population grouping	Region	Percentage prevalence
Europeans	Britain	2
	Germany	2
	Australia (1981)	2
	Australia (2002)	8
	United States	8
Native Americans	Chile Mapuche	1
	US Hispanic	17
	US Pima	50
Pacific Islanders	Nauru (1952)	0
	Nauru (2002)	41
New Guineans	Rural	0
	Urban	37
Aboriginal Australians	Traditional	0
	Westernized	23
Middle East	Yemen, traditional	4
	Yemenite Jews in Israel	13
	Lebanon, westernized	14
Black Africans	Rural Tanzania	1
	Urban South Africa	8
	United States	13
Chinese	Rural China	0
	Urban Singapore	9
	Urban Taiwan	12
	Urban Mauritius	13
Asian Indians	Rural India	0
	Urban Tanzania	11
	Urban India	12
	Urban Singapore	17
	Urban Mauritius	17
	Urban Fiji	22

Figure 1 Age-standardized prevalence of type 2 diabetes mellitus. Among the main features are the low prevalence among groups of European origin, especially those remaining in Europe; the high prevalence among Pima Indians and urban New Guineans, and among Nauruans today; and the higher prevalence in urban or westernized groups, compared with their rural or traditional counterparts. Because type 2 prevalence in a given population increases with age, it would be misleading to compare raw values of prevalence between two populations that differ in their age distribution; the raw values would be expected to differ merely as a result of the different age distributions, even if prevalences at a given age were identical between the two populations. Instead, one measures the prevalence in a population as a function of age, then calculates what the prevalence would be for that whole population if it had a certain standardized age distribution³⁰. (From refs 5, 30 and other sources.)

founded by people who undertook inter-island canoe voyages lasting several weeks. In numerous attested examples of such lengthy voyages, many or most of the canoe occupants died of starvation, and only those who were originally the fattest survived. That is why Pacific Islanders in general tend to be heavy people. Second, the Nauruans

were then set apart from most other Pacific Islanders by their extreme starvation and mortality during the Second World War, leaving the population presumably even more enriched in diabetes susceptibility genes. After the war, their new-found wealth, superabundant food and diminished need for physical activity led to exceptional obesity.

The Pimas Like other Native Americans, Arizona Pima Indians were formerly peasant farmers and hunter-gatherers who had a physically vigorous lifestyle and were at periodic risk of starvation. Their extra bout of natural selection possibly came during the late nineteenth century, when European immigrants diverted the headwaters of the rivers on which the Pimas depended for irrigation water. The result was crop failures, widespread starvation and the likely enrichment of the surviving population in thrifty genes^{8,22}.

Europe Europeans are unique among the modern world's populations in the relatively low prevalence of type 2 diabetes. Although prevalence of the disease is increasing, it is still lower than in any non-European population matched for lifestyle, even though Europeans — in Europe itself, and throughout the world — are the richest and best-fed people in the world, and the originators of the Western lifestyle. As Fig. 1 shows, even compared with the European population with the highest prevalence (white Australians, 8%), almost all other major population groupings (Native Americans, Pacific Islanders, Aboriginal Australians, East Asians and South Asians of the Indian subcontinent) include populations with much higher prevalences of 15–50%.

This uniquely low occurrence of type 2 diabetes among Europeans is curious. Several experts in the study of the disease have suggested to me informally that perhaps Europeans traditionally had little exposure to famine, so that they would have undergone little selection for a thrifty genotype. But this is not the case — there is abundant documentation of famines that have caused widespread and severe mortality in medieval and Renaissance Europe^{23–27}. So lack of exposure to famine seems unlikely to be an answer.

Instead, a more promising hypothesis is based on Europe's recent food history (see also ref. 28 for another view). The periodic widespread and prolonged famines that used to wrack Europe, like the rest of the world, disappeared between about 1650 and 1900 at different times in different parts of Europe — the late 1600s in Britain and the Netherlands, for example, and the late 1800s in southern France and southern Italy^{23–27}. With one famous exception, Europe's famines were ended by a combination of three or four factors: increasingly

efficient state intervention that rapidly redistributed any surplus grain to famine areas; increasingly efficient food transport by land and especially by sea; increasingly diversified agriculture after AD 1492, a consequence of the advent of crops, such as potatoes and corn (maize), which were brought back by European voyagers and broadened the base of European agriculture, thereby reducing the risk of starvation from failure of a single crop; and finally perhaps, Europe's reliance on 'rain agriculture', which reduced the risk of a crop failure that was too widespread to be solved by food transport within Europe, rather than (as in many populous areas outside Europe) 'irrigation agriculture'. The famous exception is, of course, the Irish potato famine of the 1840s. But this may be the exception that proves the rule. The potato famine was due to a disease of one crop in an economy that was unusual in Europe in its reliance on that single crop, and it occurred on an island governed by a state centred on another island.

A cryptic epidemic in Europe? A corollary of this view based on Europe's food history is that, several centuries before the advent of modern medicine, Europeans, like modern Nauruans, should have undergone an epidemic in type 2 diabetes that resulted from the new reliability of adequate food supplies and eliminated most diabetes-prone bearers of the thrifty genotype. However, there would have been big differences between that postulated earlier European epidemic and the well-documented modern epidemics among Nauruans and among so many other peoples today. In the modern epidemics, abundant and continually reliable food arrived suddenly, within a decade for the Nauruans and within just a month for the Yemenite Jews. The result was a sharply peaked surge in prevalence to 20–50% that occurred right under the eyes of modern diabetologists. That increase will probably wane quickly, as individuals with the thrifty genotype become eliminated by natural selection within a mere generation or two. In contrast, Europe's food abundance would have increased gradually over the course of several centuries, and the result, between the 1400s and 1700s, would have been a slow rise in type 2 prevalence long before there were diabetologists to take note.

A possible victim of this postulated cryptic epidemic of diabetes was the composer Johann Sebastian Bach (1685–1750). Bach's medical history is too poorly documented to permit certainty as to the cause of his death. Nonetheless, the corpulence of his face and hands in the sole authenticated portrait of him, the accounts of deteriorating vision in his later years, and the evident deterioration of his handwriting, possibly secondary to his failing vision, are consistent with a diagnosis of type 2 diabetes. The

disease certainly occurred in Germany during Bach's lifetime, being known as "*honigsüsse Harnruhr*" (honey-sweet urine disease)²⁹.

Tests of the hypotheses

These ideas about the evolution of type 2 diabetes can be tested. Here are some of the questions that can be asked, and predictions that can be tested.

- How much evidence is there for the postulated epidemic of diabetes in late medieval and Renaissance Europe, either in individual biographies (as that of Bach) or in contemporary medical treatises? Did the timing of the epidemic vary locally with the different times for the disappearance of famines in different parts of Europe? Was a diabetes epidemic evident after the Black Death, when human population declined much more rapidly than did food availability?

- The prevalence of type 2 diabetes in European immigrants of British and German ancestry to the United States and Australia is reported as 7–8%, much higher than the 2% for British and German people still living in Europe today under similar lifestyles. This difference is consistent with the socially stratified emigration often discussed by historians. The Europeans who stayed at home tended to be richer than those who emigrated; in the former group, the genotype predisposing the bearer to type 2 diabetes may have already been selected out by centuries of abundant food, whereas those who emigrated may have been the starvation-prone poor who still carried the thrifty genotype³. This possibility could be tested by controlled comparison of modern Europeans still living in Europe with overseas Europeans whose ancestors' country of origin and date of emigration are known. As a specific example, are overseas descendants of Irish emigrants in the 1840s more susceptible to diabetes than are Ireland's inhabitants today?

- In pre-modern times, the risk of famine was higher in the drought- and flood-prone northern areas of China than in the south. Is there a corresponding difference in predisposition to diabetes? And did China's famines during the Great Leap Forward of the late 1950s produce additional selection for thrifty genes?

- Were the risks of famine, and so selection for thrifty genes, higher in societies that depended heavily on fishing and hunting-and-gathering (Nauruans, Pimas and Aboriginal Australians, for instance) than in farming societies in which food storage had become the norm (Europe and China)?

- Are genetic susceptibility factors for type 2 diabetes especially evident in Nauruans, Yemenite Jews and other populations showing recent surges of the disease?

- Anecdotal accounts suggesting that there were other modern epidemics of diabetes deserve investigation. For instance, a colleague whose grandfather was from northern

Iran recounts that improved food transport there in the early 1900s reduced the frequency of starvation and triggered a diabetes epidemic, especially among rich people — hence the local term 'the rich man's disease'.

- Medical geneticists seek to identify and preemptively counsel people carrying susceptibility factors for specific diseases. Can Third World populations, whose history of marked swings between food and starvation puts them at risk of diabetes with the spread of colonialization, be likewise identified and preemptively counselled by public-health officials? That might spare them the fate of the Nauruans, Arizona Pima Indians and Wani-gela. The evolutionary history and geography of type 2 diabetes would then have provided us not only with a double puzzle, but also with insights that could potentially help save millions of people from premature death. ■

Jared Diamond is in the Departments of Geography and Environmental Health Sciences, University of California, Los Angeles, 1255 Bunche Hall, Box 951524, California 90095-1524, USA.

e-mail: jdiamond@geog.ucla.edu

- Levitt, N. *et al.* *Diabetes Care* **16**, 601–607 (1993).
- Zimmet, P., Alberti, K. & Shaw, J. *Nature* **414**, 782–787 (2001).
- Zimmet, P. in *The Medical Challenge: Complex Traits* (eds Fischer, E. & Moller, G.) 55–110 (Piper, Munich, 1997).
- Zimmet, P. *J. Intern. Med.* **247**, 301–310 (2000).
- King, H., Aubert, R. & Herman, W. *Diabetes Care* **21**, 1414–1431 (1998).
- Dowse, G., Zimmet, P., Finch, C. & Collins, V. *Am. J. Epidemiol.* **133**, 1093–1104 (1991).
- Rubinstein, H. & Zimmet, P. *Phosphate, Wealth, and Health in Nauru: A Study of Lifestyle Change* (Brolga, Gundaroo, 1993).
- Knowler, W., Pettitt, D., Saad, N. & Bennett, P. *Diabetes Metab. Rev.* **6**, 1–27 (1990).
- Dowse, G. *et al.* *Med. J. Aust.* **160**, 767–774 (1994).
- Gardner, L. *et al.* *Diabetes* **33**, 86–92 (1984).
- Elbein, S., Chiu, K. & Permutt, M. in *The Genetic Basis of Common Diseases* 2nd edn (eds King, R., Rotter, J. & Motulsky, A.) 457–480 (Oxford Univ. Press, 2002).
- Raffel, L. & Rotter, J. in *Principles and Practice of Medical Genetics* Vol. 2, 4th edn (eds Rimoin, D., Connor, J. & Pyeritz, R.) 2231–2276 (Churchill Livingstone, London, 2001).
- Hales, C. & Barker, D. *Diabetologia* **35**, 595–601 (1992).
- Zimmet, P. *Diabetes Care* **15**, 232–252 (1992).
- Campbell, G. S. *Afr. Med. J.* **37**, 1195–1207 (1963).
- Coleman, D. *Science* **203**, 663–665 (1979).
- Dowse, G. K. *et al.* *Diabetes* **39**, 390–396 (1990).
- Diamond, J. *Nature* **410**, 521 (2001).
- Neel, J. *Am. J. Hum. Genet.* **14**, 353–362 (1962).
- Neel, J. in *The Genetics of Diabetes Mellitus* (eds Kobberling, J. & Tattersall, J.) 283–293 (Academic, New York, 1982).
- Haines, H., Hackel, D. & Schmidt-Nielsen, K. *Am. J. Physiol.* **208**, 297–300 (1965).
- Russell, F. *The Pima Indians* (Univ. Arizona Press, Tucson, 1975).
- Appleby, A. *Famine in Tudor and Stuart England* (Stanford Univ. Press, 1978).
- Jordan, W. *The Great Famine: Northern Europe in the Early Fourteenth Century* (Princeton Univ. Press, 1996).
- Monahan, W. *Year of Sorrows: The Great Famine of 1709 in Lyon* (Ohio State Univ. Press, Columbus, 1993).
- Post, J. *Food Shortage, Climatic Variability and Epidemic Disease in Preindustrial Europe* (Cornell Univ. Press, Ithaca, 1985).
- Rotberg, R. & Rabd, T. (eds) *Hunger and History: The Impact of Changing Food Production and Consumption Patterns on Society* (Cambridge Univ. Press, 1983).
- McMichael, A. *Human Frontiers, Environments and Disease: Past Patterns on Certain Futures* (Cambridge Univ. Press, 2001).
- Kranemann, D. *Bach Jahrbuch* 53–64 (1990).
- King, H. & Rewers, M. *Diabetes Care* **16**, 157–177 (1993).

It is a pleasure to acknowledge my debt for fruitful discussions about diabetes with Paul Zimmet and Jerome Rotter, and about European food history with Jack Goldstone and William Jordan.

Ice-age steppe vegetation in east Beringia

Tiny plant fossils indicate how this frozen region once sustained huge herds of mammals.

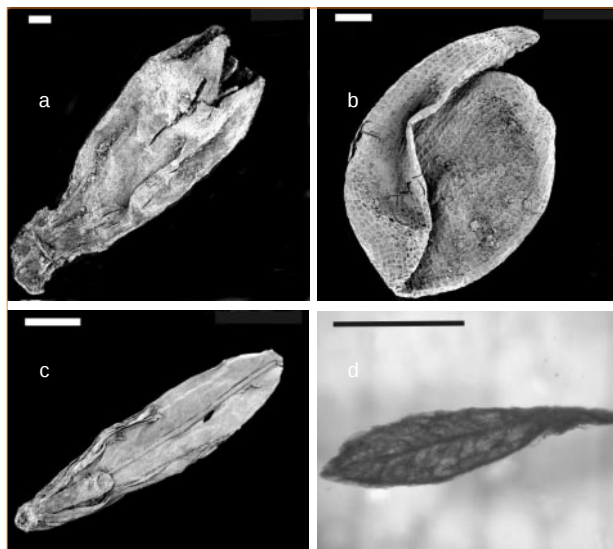
The landmass known as Beringia is an extensive region that existed during the Pleistocene epoch and included the land bridge between present-day Siberia and Alaska, now submerged beneath the Bering Strait. It must have been covered with vegetation even during the coldest part of the most recent ice age (some 24,000 years ago) because it supported large populations of woolly mammoth, horses, bison and other mammals during a time of extensive Northern Hemisphere glaciation, although the nature of this vegetation has not been determined^{1–3}. Here we report the discovery of macrofossils of prairie sage (*Artemisia frigida*), bunch-grasses and forbs that are representative of ice-age steppe vegetation associated with Pleistocene mammals in eastern Beringia. This vegetation was unlike that found in modern Arctic tundra, which can sustain relatively few mammals, but was instead a productive ecosystem of dry grassland that resembled extant subarctic steppe communities^{4,5}.

Spectra of fossil pollens indicate that there was a high proportion of sage (*Artemisia*) and grass (Poaceae) in eastern Beringia⁶, suggesting that this region might have contained an arid but productive, grass-dominated ecosystem. This system, called a 'mammoth-steppe'², would have sustained mammalian herds all year round. The fossil pollens have alternatively been interpreted as representing a discontinuous herb-tundra ecosystem³. Limited taxonomic resolution of dominant pollen types precludes the clear differentiation of steppe from tundra vegetation and calls for more conclusive evidence^{1,7,8}.

We analysed plant macrofossils (Fig. 1) from three late-Pleistocene sites in Yukon Territory, Canada. These were recovered from a rodent nest, radiocarbon-dated at about 24,000 years before present (yr BP), in Quartz Creek in western central Yukon⁹; from the stomach contents of a permafrost-frozen carcass of an extinct horse (*Equus lambei*, 26,280 ± 210 yr BP)⁹; from an alluvial peat (25,700 ± 400 yr BP) associated with woolly-mammoth remains found in Last Chance Creek, also in western central Yukon; and from a 3.5-metre-thick sediment sequence (18,800–16,400 yr BP), from Bluefish River exposure in northern Yukon. These assemblages reflect the local vegetation that would have been associated with fossil mammals in easternmost Beringia.

Our macrofossils are dominated by sage, bunch-grasses and grass-like plants, and include other vegetation that would have flourished in cold, dry steppe conditions.

Figure 1 Late-Pleistocene plant macrofossils from Yukon Territory, Canada. **a**, Sage (*Artemisia*) flower; **b**, mustard (*Draba*) seed; **c**, wild-rye grass (*Elymus*) seed; **d**, prairie sage (*Artemisia frigida*) leaf. Fossils were identified by comparison with herbarium specimens. Scale bars: **a**, **b**, 0.1 mm; **c**, 1 mm; **d**, 2 mm.



Abundant sage (*Artemisia frigida*) leaves, flowers from *Artemisia* sp., and seeds of bluegrass (*Poa*), wild-rye grass (*Elymus*), sedge (*Carex*) and rushes (*Juncus/Luzula*) support this interpretation. Seeds of cinquefoil (*Potentilla*), goosefoot (*Chenopodium*), buttercup (*Ranunculus*), mustard (*Draba*), poppy (*Papaver*), fairy-candelabra (*Androsace septentrionalis*), chickweed (*Cerastium*) and campion (*Silene*) are indicative of diverse forbs growing on dry, open, disturbed ground, possibly among predominantly arid-steppe vegetation^{4,5}. Such an assemblage has no modern analogue in Arctic tundra³. Local habitat diversity is indicated by sedge and moss peat from deposits that were formed in low-lying wet areas.

The absence of sage and bunch-grass macrofossils from central Beringia suggests that steppe flora did not cover all of Beringia^{1,7,8}. Instead, it is likely that local ecological mosaics existed, on the basis of evidence of wet tundra in poorly drained lowlands⁷ and of dry, herb-and-sedge (*Kobresia* sp.) upland tundra that developed under the influence of loess deposition⁸. Macrofossil information indicates that local factors, including soil moisture, loess deposition, slope and altitude, were important in determining regional Beringian vegetation¹. Comparison of the features of these macrofossils indicates that there were likely to have been marked ecological differences between the central and easternmost regions of Beringia during the late Pleistocene.

In Alaska and Yukon today, *A. frigida*, bunch-grasses and forb associations are restricted to small, isolated steppe communities on arid, south-facing slopes with well-

drained soils, deep active layers and high net insolation^{9,10}. Regional full-glacial climatic aridity⁶ provided similar conditions that favoured the widespread establishment of upland steppe flora in eastern Beringia. We conclude that our macrofossils are evidence of the co-occurrence of local steppe vegetation and grazing mammals within the Yukon interior, which characterized the supposed mammoth-steppe ecosystem².

Grant D. Zazula*, **Duane G. Froese†**, **Charles E. Schweger‡**, **Rolf W. Mathewes***, **Alwynne B. Beaudoin§**, **Alice M. Telka||**, **C. Richard Harington¶**, **John A. Westgate#**

Departments of *Biological Sciences and †Earth Sciences, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada
e-mail: gdzazula@sfu.ca

‡Department of Anthropology, University of Alberta, Edmonton, Alberta T6G 2H4, Canada

§Provincial Museum of Alberta, Edmonton, Alberta T5N 0M6, Canada

||PALEOTEC Services, West Ottawa, Ontario K1R 5K2, Canada

¶Canadian Museum of Nature, PO Box 3443, Ottawa, Ontario K1P 6P4, Canada

#Department of Geology, University of Toronto, Toronto, Ontario M5S 3B1, Canada

1. Birks, H. H. & Birks, H. J. B. *J. Biogeogr.* **27**, 31–35 (2000).

2. Guthrie, R. D. *Frozen Fauna of the Mammoth Steppe* (Univ. Chicago Press, Chicago, 1990).

3. Cwynar, L. C. & Ritchie, J. A. *Science* **208**, 1375–1377 (1980).

4. Edwards, M. E. & Armbruster, W. S. *Arch. Alp. Res.* **21**, 296–304 (1989).

5. Vetter, M. A. *Arctic* **53**, 165–173 (2000).

6. Anderson, P. M. & Brubaker, L. B. *Q. Sci. Rev.* **13**, 71–92 (1994).

7. Elias, S. A. *et al. Nature* **382**, 60–63 (1996).

8. Goetcheus, V. G. & Birks, H. H. *Q. Sci. Rev.* **20**, 135–147 (2001).

9. Froese, D. *et al. Q. Sci. Rev.* **21**, 2137–2142 (2002).

10. Harington, C. R. & Eggleston-Stott, M. *Curr. Res. Pleist.* **13**, 105–106 (1996).

Competing financial interests: declared none.

Insect behaviour

Motion camouflage in dragonflies

Most animals can skilfully conceal themselves when stationary¹, but they may become apparent as soon as they move. Here we use stereo cameras to reconstruct the movements in three dimensions of dragonflies (*Hemianax papuensis*), and show that these insects actively use motion camouflage to disguise themselves as stationary during territorial aerial manoeuvres. Deployment of this sophisticated technique by the oldest airborne predator tricks the victim's retina into perceiving the stalker as stationary even while it darts about in pursuit.

Optic flow — the apparent movement of objects as perceived using the retina — is a primary cue for detecting predators and prey. Predatory animals attacking stationary prey generally attempt to conceal their presence by approaching very slowly, sometimes relying on a camouflaging background. Active motion camouflage has been proposed as a strategy by which a predator can conceal its movements while shadowing or attacking highly manoeuvrable prey².

Motion camouflage can be achieved if one animal (the shadower) moves in such a way as to produce the same image motion on the retina of another animal (the shadowee) as would a stationary object in the environment. We reconstructed 15 three-dimensional flight trajectories of interactions between conspecific dragonflies, of which six showed clear evidence of active motion camouflage.

Figure 1a shows an interaction between two male dragonflies: the lines connecting the heads of the shadower (blue dots) and shadowee (red dots) intersect in a small region behind the shadower. The centre of the shaded sphere represents the best estimate of the common point of intersection, and the radius of the sphere represents the uncertainty of this estimate.

If motion camouflage were perfect, all lines would intersect at a single point, and the sphere of uncertainty would have zero radius. In this example, the radius of the intersection sphere is 11.3 mm. Given the ± 10 -mm accuracy in measuring position at a filming distance of 1.5 m, we may say that the lines intersect almost at a common point. This indicates that the shadower is emulating the trajectory of the projection of the fixed intersection point on the shadowee's retina. In other words, as far as the shadowee is concerned, the shadower is indistinguishable from a stationary object positioned at the intersection point.

The effectiveness of the strategy of motion camouflage is illustrated by a comparison of the apparent motion of the shadower on the

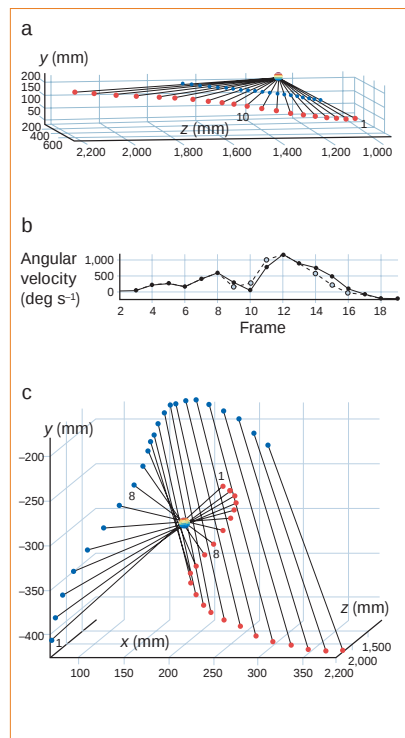


Figure 1 Examples of active motion camouflage. **a**, Three-dimensional reconstruction of a territorial interaction involving two male dragonflies. The reconstruction was achieved by using calibrated⁶ and synchronized (to within 10 μ s) stereo cameras. Positions of the heads of the shadower (blue dots) and shadowee (red dots) are connected by lines at corresponding 20-ms steps. The connecting lines converge behind the shadower; the point of convergence indicates the apparent position of the shadower to the shadowee. The shaded sphere (radius, 11.3 mm) represents the uncertainty of the 'best' estimate of the intersection point. **b**, Angular-velocity profile produced by the shadowing dragonfly in the shadowee's eye (filled circles) compared with that produced by a virtual stationary object at the intersection point (hollow circles). **c**, Another example of motion camouflage. The connecting lines converge in front of the shadower during frames 1–8, after which the shadower begins to imitate an infinitely distant object. The radius of the intersection sphere is 8.5 mm.

shadowee's retina with that produced by a virtual, stationary object positioned at the estimated intersection point (Fig. 1b). The two angular-velocity profiles are very similar, indicating that the strategy is likely to be effective in concealing the motion of the shadower.

In other male–male interactions, we observed variations in motion camouflage (Fig. 1c). During the first eight frames of filming, the lines connecting the two insects intersect at a small region between them; the shadower is therefore emulating a fixed object in front of it. Here the shadowing animal is not flying towards the shadowee, but is moving in the opposite direction. This is evidence that the flight pattern of the shadower is motivated by motion camouflage and is not an artefact or by-product of chasing *per se*^{2,3}. After frame nine, the intersection point shifts to

infinity — that is, the shadower imitates an object at infinity. This example reveals that a shadower can combine two different types of motion camouflage in a single episode.

Active motion camouflage requires the shadower to move in such a way that it imitates the trajectory of a fixed object on the retina of the shadowee by precise flight control and positional sensing⁴, although exactly how the shadower achieves this is unclear². Given the high sensitivity of the insect visual system to movement⁵, motion camouflage is likely to be the key to the success of such manoeuvres.

Akiko Mizutani*, **Javaan S. Chahl*†**, **Mandyam V. Srinivasan***

*Centre for Visual Science, Research School of Biological Sciences, Australian National University, PO Box 475, Canberra, ACT 2601, Australia
e-mail: akiko@biorobotics.anu.edu.au

†Weapons Systems Division, Defence Science and Technology Organisation, Edinburgh, South Australia 5111, Australia

- Curio, E. *The Ethology of Predation* (Springer, Berlin, 1976).
- Srinivasan, M. V. & Davey, M. *Proc. R. Soc. Lond. B* **259**, 19–25 (1995).
- Collett, T. S. & Land, M. F. *J. Comp. Physiol.* **99**, 1–66 (1975).
- US Office of Naval Research *Insect Vision Based Seekers and Controllers for Guided Weapons* (contract N68936-1-2-001).
- Srinivasan, M. V. in *Visual Motion and its Role in the Stabilization of Gaze* (eds Miles, F. A. & Wallman, J.) 139–155 (Elsevier, Amsterdam, 1993).
- Tsai, R. Y. *IEEE J. Robot. Automat.* **3**, 323–344 (1987).

Competing financial interests: declared none.

Electromagnetic waves

Negative refraction by photonic crystals

Materials that can bend light in the opposite direction to normal ('left-handed' materials) reverse the way in which refraction usually works — this negative refractive index is due to simultaneously negative permeability and permittivity^{1–3}. Here we demonstrate negative refraction of electromagnetic waves in a two-dimensional dielectric photonic crystal that has a periodically modulated positive permittivity and a permeability of unity^{4–6}. This experimental verification of negative refraction is a step towards the realization of a 'superlens' that will be able to focus features smaller than the wavelength of light.

Our structure consists of a square array of alumina rods in air; the calculated band diagram for this structure is shown in Fig. 1a. To obtain negative refraction, equal-frequency surfaces are needed for the photonic crystal that are both convex and larger than those for air (Fig. 1b). Note that conservation of the surface-parallel wave vector gives the direction of the refracted waves inside the photonic crystal (Fig. 1b).

We made transmission measurements

Sexual contacts and epidemic thresholds

Distributions of the number of sexual partners reported in surveys show a pronounced skew, with most people having had one or no partners in the past year and a small fraction having had many^{1,2}. Liljeros and colleagues³ infer from the results of a Swedish survey that there is a “scale-free” population distribution of sexual contacts, consistent with a preferential-attachment model^{3,4}, in which “the rich get richer” and epidemics are driven by extremely promiscuous individuals. Here we reanalyse the data from Sweden and from other countries, using more appropriate statistical tools. Our findings support the conventional wisdom that epidemic thresholds exist in these populations, and indicate that current public-health strategies to reduce the spread of HIV and other sexually transmitted infections do not need to be radically refocused.

An important epidemiological question for highly skewed partnership distributions is whether their variance is finite^{5,6}. As the reproduction number of a pathogen increases linearly with the variance in the level of sexual activity^{7,8}, populations with infinite variance lack epidemic thresholds. In these populations, a sexually transmitted infection could persist regardless of its transmissibility, and interventions such as vaccines or barrier contraceptives would be ineffective for eradicating it.

Liljeros *et al.* estimate the scaling exponent for Sweden by fitting a line to the apparently linear region of the upper tail of the logged sexual-contact distribution. This approach is not statistically appropriate, for several reasons⁹. Inference on the basis of the distribution's extreme upper tail yields wildly increasing confidence intervals because there is little empirical information in this region (Fig. 1). Estimates based on partial lifetime contacts (as in Fig. 2b of ref. 3) are a source of further difficulties, including temporal confounding and censoring.

We take a more principled statistical approach, using a stochastic mechanism for the underlying preferential-attachment process. This yields a Yule distribution¹⁰, and infinite variance when the single scaling parameter $\rho \leq 3$. (Details of the Yule model and the statistical estimation procedure are presented elsewhere⁹.) We generalize the simple Yule model to allow for a mixture of distributions — a process for the lower tail, and a preferential-attachment-type process for the upper tail. Assuming the Yule form, we can estimate the model parameter using maximum

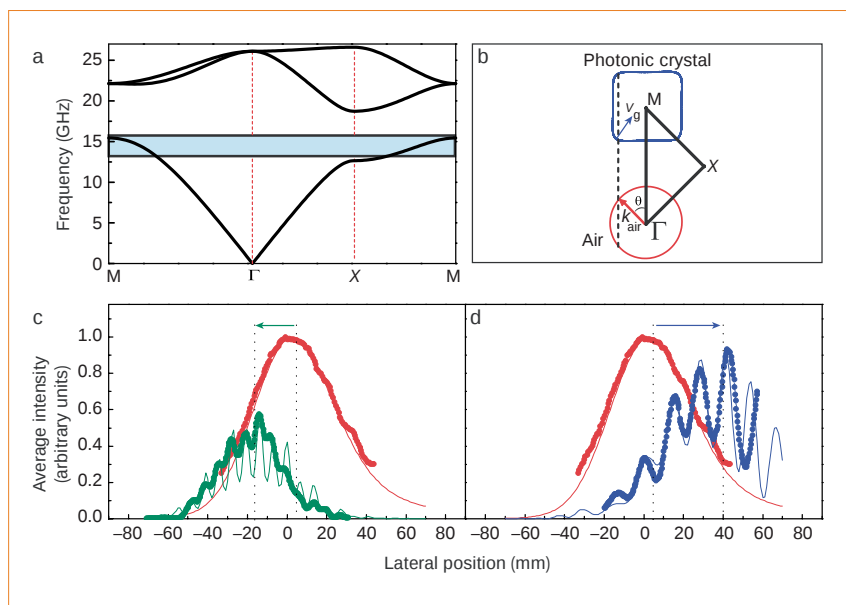


Figure 1 Comparison of positive and negative refraction. **a**, Band diagram for our structure for transverse magnetic polarization. Shading denotes the frequency range in which negative refraction occurs. **b**, Equal-frequency contours in k space of air and of the photonic crystal at 13.7 GHz. θ is the angle of incidence from the air to the crystal; v_g is the group velocity inside the photonic crystal. **c**, Negative refraction. Average intensities were calculated at the second interface with the photonic crystal (green) and at the first interface without the photonic crystal (red), and corresponding power distributions were measured. **d**, Positive refraction. Average intensities were calculated at the second interface with a slab containing polystyrene pellets (blue) and at the first interface without the slab (red), and the corresponding power distributions were measured. Arrows in **c** and **d** indicate the refracted beam's direction of divergence from the incident beam.

to confirm our structure's predicted negative refraction, using the interfaces of the photonic crystal in the Γ – M direction. The electric field was kept parallel to the rods for all measurements and calculations; the horn antenna was orientated so that the incident waves make an angle of 45° with the normal of the Γ – M interface. Our structure exhibits the maximum angular range of negative refraction at an operating frequency of 13.7 gigahertz (GHz). In simulations using the finite-difference time-domain method (FDTD), the incident gaussian beam width was set at 6 cm, which is equal to the width of the horn antenna.

The centre of the outgoing gaussian beam is shifted to the left of the centre of the incident gaussian beam (Fig. 1c), which corresponds to negative refraction⁷. The negative index of refraction was determined to be -1.94 , which is close to the theoretical value of -2.06 calculated by the FDTD method. For comparison, we repeated the measurements and simulations with a slab containing only polystyrene pellets, which has a refractive index of 1.46, and found the refracted beam to be on the right of the incident beam, corresponding to a positive index of 1.52 (Fig. 1d).

The advantage of negative refraction in the valence band is that there is no Bragg reflection; such reflections occur in higher bands of the photonic crystal, and we have a well-defined, single-beam propagation that is negatively refracted. Another advantage of operating in the valence band is that the

transmission efficiency at this frequency is 63%, which is almost three orders of magnitude larger than the typical transmission efficiency in a left-handed material^{2,3}. The negative-refraction effect that we describe depends only on the refractive index of the dielectric material and on the geometric factors used in two-dimensional photonic crystals. This effect can therefore also be observed at optical wavelengths, at which it is possible to obtain similar refractive indices by using transparent semiconductors.

Ertugrul Cubukcu*, **Koray Aydin***,
Ekmel Ozbay*, **Stavroula Foteinopoulou†**,
Costas M. Soukoulis†‡

*Department of Physics, Bilkent University,
Bilkent, 06533 Ankara, Turkey

e-mail: cubukcu@fen.bilkent.edu.tr

†Ames Laboratory, US Department of Energy,
and Department of Physics and Astronomy,
Iowa State University, Ames, Iowa 50011, USA

‡Research Center of Crete and Department of
Materials Science and Technology, University of
Crete, Heraklion, 71100 Crete, Greece

1. Pendry, J. B., Holden, A. J., Robins, D. J. & Stewart, W. J. *IEEE Trans. Microwave Theory Tech.* **47**, 2075–2084 (1999).
2. Shelby, R. A., Smith, D. R. & Schultz, S. *Science* **292**, 77–79 (2001).
3. Bayindir, M., Aydin, K., Ozbay, E., Markos, P. & Soukoulis, C. M. *Appl. Phys. Lett.* **81**, 120–122 (2002).
4. Notomi, M. *Phys. Rev. B* **62**, 10696–10705 (2000).
5. Luo, C., Johnson, S. G., Joannopoulos, J. D. & Pendry, J. B. *Phys. Rev. B* **65**, 201104(R) (2002).
6. Foteinopoulou, S., Economou, E. N. & Soukoulis, C. M. *Phys. Rev. Lett.* **90**, 107402 (2003).
7. Kong, J. A., Wu, B. I. & Zhang, Y. *Appl. Phys. Lett.* **80**, 2084–2086 (2002).

Competing financial interests: declared none.

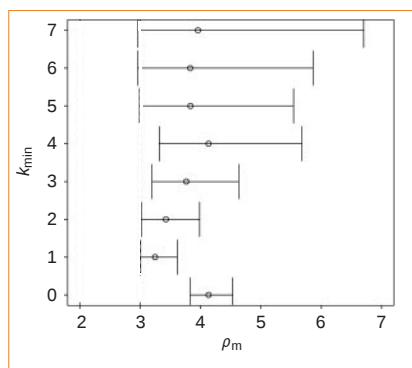


Figure 1 Interval estimates of the scaling parameter, ρ , for the generalized Yule probability mass function for Swedish males and females. Data show the sensitivity of the 95% bootstrap confidence intervals to the upper tail of the partnership distribution, defined as $k > k_{min}$, where k is the number of sexual partners in the previous 12 months. For both males and females, the best-fitting model has $k_{min} = 1$. By contrast, Liljeros *et al.*³ used $k_{min} = 5$ for males and $k_{min} = 4$ for females. For these values, our estimates of parameter uncertainty are six times greater than the intervals reported by Liljeros *et al.*³.

likelihood, and base our model selection on the Bayesian information criterion⁹.

The 95% bootstrap confidence-interval estimates of the scaling parameters fall above the infinite-variance range (females, 3.60–5.21; males, 3.01–3.63). Our results are similar when we analyse two further data sets from Uganda and the United States (both based on representative samples with substantially higher response rates than in the Swedish survey). Thus, there seems to be consistent evidence that sexual-contact distributions are characterized by finite variance⁹.

Our findings suggest that sexual-contact distributions, although strongly right-skewed, are characterized by finite variance. This implies that interventions aimed at reducing transmissibility have the potential to reduce the reproductive number of sexually transmitted infections below the epidemic threshold.

To justify using “radically different” prevention strategies for sexually transmitted infections, strong evidence is needed that current strategies will be ineffective. Results based on unverified mathematical assumptions about network structure do not come close to meeting this standard¹¹. Such data provide no new insight for epidemiology^{7,8}, and conclusions drawn from them could jeopardize campaigns to eliminate sexually transmitted infections worldwide.

It has been suggested in the media that ‘eliminating’ highly promiscuous nodes could be more effective than reducing their probability of transmission, but this overlooks the remarkable progress in preventing mother-to-child infection, the difficulty in identifying highly active individuals, and the growing evidence of the importance of other behavioural patterns, such as differential

assortative mixing¹ and concurrency^{12,13}. The high stakes in the battle against HIV and AIDS call for a broad perspective.

James Holland Jones, Mark S. Handcock
Center for Statistics and the Social Sciences, and
Center for AIDS and Sexually Transmitted
Diseases, University of Washington, Seattle,
Washington 98195, USA
e-mail: jameshj@stat.washington.edu

1. Laumann, E. O., Gagnon, J. H., Michael, T. & Michaels, S. *The Social Organization of Sexuality: Sexual Practices in the United States* (Univ. Chicago Press, Chicago, 1994).
2. Lewin, B. (ed.) *Sex in Sweden* (Nat'l Inst. Pub. Health, Stockholm, 1996).
3. Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E. & Åberg, Y. *Nature* **411**, 907–908 (2001).
4. Barabási, A. L. & Albert, R. *Science* **286**, 509–512 (1999).
5. May, R. M. & Lloyd, A. L. *Phys. Rev. E* **64**, 066112 (2001).
6. Johnson, N. L., Kotz, S. & Kemp, A. W. *Univariate Discrete Distributions* (Wiley, New York, 1992).
7. Anderson, R. M., Medley, G. F., May, R. M. & Johnson, A. M. *IMA J. Math. Appl. Med. Biol.* **3**, 229–263 (1986).
8. Anderson, R. M. & May, R. M. *Infectious Diseases of Humans: Dynamics and Control* (Oxford Univ. Press, Oxford, 1991).
9. Jones, J. H. & Handcock, M. S. *Proc. R. Soc. Lond. B* published online (doi:10.1098/rspb.2003.2369).
10. Simon, H. *Biometrika* **42**, 435–440 (1955).
11. Pastor-Satorras, R. & Vespignani, A. *Phys. Rev. E* **65**, 035108 (2002).
12. Morris, M. & Kretzschmar, M. *AIDS* **11**, 641–648 (1997).
13. Adimora, A. A. *et al. Epidemiology* **13**, 320–327 (2002).

Liljeros et al. reply — Jones and Handcock have reanalysed two of our four data sets of the number of different sexual partners for Swedish men and women. Specifically, they analyse the probability-distribution function $p(k)$, which yields less reliable information than the cumulative distribution,

$$P(\geq k) = \sum_{k' \geq k} p(k')$$

that we used in our analysis, and they propose that the tails of $p(k)$ decay with a power-law exponent $\rho > \rho_c = 3$, which is larger than the threshold value above which the variance of the distribution is finite. They also question whether public-health strategies need to be refocused in view of their inferred existence of epidemic thresholds.

We argue that these claims are misleading. Jones and Handcock's point that the tail exponent $\rho > \rho_c = 3$, the threshold for finite variance, is not a new one. Their estimates, in fact, agree with ours¹, which indicate that the tails of the distributions of the number of sexual partners decay asymptotically as power laws with $\rho_c > 3$: the apparent discrepancy arises because Jones and Handcock use the probability-distribution function, $p(k)$, whereas we studied the cumulative distribution, $P(\geq k)$. The exponent ρ is related to the exponent α by the equation $\rho = \alpha + 1$; hence, our values of α are the same as those that Jones and Handcock derive for ρ .

Jones and Handcock ignore two of our data sets¹, in which we analyse the number of lifetime sexual partners. This is important in view of the duration of the infectious period of some sexually transmitted pathogens such

as HIV. When we re-analyse the data sets for the number of lifetime partners using the Yule distribution, we obtain exactly the same result as for a power-law fit because the Yule distribution closely resembles a power law in the interval analysed (further details are available from the authors).

Jones and Handcock argue that if the variance of the distribution of sexual partners is finite, then interventions aimed at reducing disease transmissibility have the potential to eradicate sexually transmitted infections. The two data sets that they analysed pertain to the number of different sexual partners during the previous year, rendering any discussion about the existence of epidemic thresholds speculative (because other factors besides the distribution of the number of partners in the previous year influence the existence of an epidemic threshold for infinitely large networks^{2–5}). For finite networks, the variance of these distributions must be finite, as we are analysing finite populations of individuals who are sexually active for finite periods of time⁶. We did not, therefore, raise the issue of infinitely low thresholds in our communication¹.

We contend that Jones and Handcock's conclusion is premature, given that contagious processes differ fundamentally for scale-free and single-scale networks^{4,5} and that, if a threshold were to exist, it would be very low, as the variance of the distribution of the number of sexual partners is much larger than the mean^{5,7,8}.

Jones and Handcock claim that there is no need for a radical refocusing of current public-health strategies⁹. We agree, as targeting high-risk sexual behaviour and carrying out contact-tracing routines — the strategies best suited to scale-free networks — are well established practices in many countries.

Fredrik Liljeros*†, Christofer R. Edling†, H. Eugene Stanley‡, Y. Åberg†, Luis A. N. Amaral§

*Swedish Institute for Infectious Disease Control, 171 82 Solna, Sweden

†Department of Sociology, Stockholm University, 106 91 Stockholm, Sweden
e-mail: liljeros@sociology.su.se

‡Centre for Polymer Studies, Department of Physics, Boston University, Boston, Massachusetts 02215, USA

§Department of Chemical Engineering, Northwestern University, Evanston, Illinois 60208, USA

1. Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E. & Åberg, Y. *Nature* **411**, 907–908 (2001).
2. Liljeros, F., Edling, C. R. & Amaral, L. A. N. *Microbes Infect.* **5**, 189–196 (2003).
3. Pastor-Satorras, R. & Vespignani, A. *Phys. Rev. E* **65**, 036104 (2002).
4. Newman, M. E. J. *Phys. Rev. E* **66**, 016128 (2002).
5. May, R. M. & Lloyd, A. L. *Phys. Rev. E* **64**, 066112 (2001).
6. Amaral, L. A. N., Scala, A., Barthélemy, M. & Stanley, H. E. *Proc. Natl Acad. Sci. USA* **97**, 11149–11152 (2000).
7. Brewer, D. D. *et al. Proc. Natl Acad. Sci. USA* **97**, 12385–12388 (2000).
8. Stirling, A. C., Stanley, E. A., Hyman, J. M., Layne, S. P. & Qualls, C. *Proc. Natl Acad. Sci. USA* **86**, 4793–4797 (1989).
9. Hethcote, H. & Yorke, J. A. *Gonorrhea Transmission Dynamics and Control* (Springer, Berlin, 1984).

Three modes of synaptic vesicular recycling revealed by single-vesicle imaging

Sunil P. Gandhi*† & Charles F. Stevens†

*Neurosciences Program, University of California, San Diego, La Jolla, California 92093, USA

†Howard Hughes Medical Institute, and Molecular Neurobiology Laboratory, The Salk Institute, La Jolla, California 92037, USA

Synapses recycle their spent vesicles in order to keep up with on-going neurotransmitter release. To investigate vesicle recycling in the small synapses of hippocampal neurons, we have used an optical recording method that permits us to resolve single-vesicle events. Here we show that an exocytic event can terminate with three modes of vesicle retrieval: a fast (400–860 ms) ‘kiss-and-run’ mode that has a selective fusion pore; a slow (8–21 s) ‘compensatory’ mode; and a ‘stranded’ mode of recycling, in which a vesicle is left on the cell surface until a nerve impulse triggers its retrieval. We have also observed that, in response to a nerve impulse, synapses with low release probability primarily use the kiss-and-run mode, whereas high release probability terminals predominantly use the compensatory mode of vesicle retrieval.

As synapses draw on a limited pool of releasable vesicles^{1,2}, how fast this pool can be replenished sets the maximal steady-state rate for synaptic transmission. One of the major rate-limiting steps in replenishment is the retrieval of used vesicles from the cell surface. Despite extensive study^{3–17}, the means of vesicle retrieval remains controversial for the small synapses in brain. Some of this controversy arises because it has not been possible to measure the endocytic rate of individual vesicles at these synapses with good time resolution.

In this study we use a pH sensor of exo- and endocytosis—synaptophysin, developed by Meisenbock and colleagues¹⁸—that provides a fluorescence signal when the lumen of the vesicle is exposed to the external medium (see Fig. 1a). By optimizing the imaging conditions, we can detect single-vesicle recycling events with this sensor.

We overexpressed the synaptophysin protein in 2–3-week-old cultured hippocampal pyramidal neurons using the Sindbis viral delivery method (see Methods). Figure 1b shows an image of an infected culture preparation in which transmitted light differential interference contrast (DIC) and fluorescence images are superimposed. The synaptophysin fluorescence appears largely punctate in axons, revealing the sites of putative synapses.

To identify synaptic vesicle cycling sites we imaged axon segments while delivering a 20-action-potential probe (at 20 Hz) with field stimulation. Image frames were gathered at a 10-Hz sample rate. Subtracting the frame gathered before stimulation from the one immediately after, we obtained a difference image (Fig. 1c). After centring square regions of interest (ROIs) over the difference image pixels showing a positive change, we integrated the pixel values in the ROI for each frame to assemble the fluorescence traces in Fig. 1d. Increases in fluorescence were confined to the stimulation period (Fig. 1d) and on average (Fig. 1d, bottom trace) took more than 10 s to return to baseline.

It is evident from the traces in Fig. 1d that the background noise would make single-vesicle resolution very difficult. Fortunately, we discovered (see Supplementary Information) that pre-photobleaching the terminals could attenuate the background fluorescence up to 90%, while leaving the amplitude of the evoked signal change unaffected. As a consequence of pre-photobleaching, the estimated single-vesicle signal-to-noise ratio approached 2:1.

Detection of single-vesicle exocytic events

Armed with the improvement provided by photobleaching, we could extend the per-vesicle signal-to-noise ratio to 4:1 with some time averaging. The procedure that we used was to estimate a

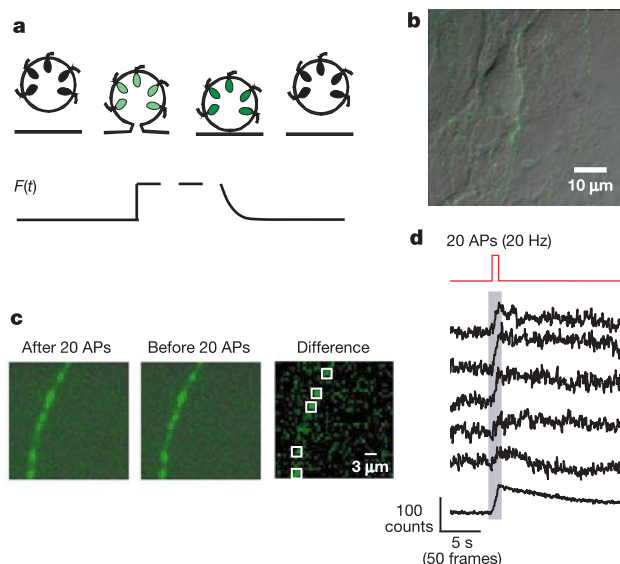


Figure 1 Imaging synaptic vesicle recycling by sensing vesicular pH. **a**, A pH-sensitive mutant of GFP¹⁸ is anchored in the vesicle lumen. At neutral pH the GFP is fluorescent but at low pH it is quenched³⁴ ($pK_a = 7.1$). A synaptic vesicle is tonically acidic^{11,35,36}, so the sensor is quenched. On exocytosis, accumulated H^+ leave the vesicle and the sensor is unquenched. Fluorescence persists until the vesicle is retrieved and the H^+ pump can acidify the lumen³⁷. **b**, Superposition of fluorescence and DIC images of cultured hippocampal neurons after viral delivery of sensor. Fluorescence largely labels axons with dim puncta. **c**, Vesicle release sites are identified by the difference image (calculated by subtracting the frame recorded at the start of the 20-action-potential (AP) stimulation from the one at the end). ROIs (white boxes, $2.6 \mu m^2$) are centred on spots of evoked fluorescence along the axon. **d**, Time course of fluorescence for individual ROIs in **c** (five upper traces), and average of 50 traces (bottom). Signal increases only during stimulation and decays slowly ($\tau \approx 10$ s).

response by averaging fluorescence intensity over a synapse for a few hundred milliseconds (usually 500 ms or five frames) after stimulation, and by subtracting from this average a baseline estimate made by averaging, for the same synapse, the 2.5 s (25 frames) preceding stimulation (see Methods). The calculated fluorescence changes (ΔF values) could be compiled into an amplitude histogram.

As a first step in resolving single release events, we characterized background noise by measuring the distribution of ΔF values elicited by one action potential in low calcium concentration bathing solution (0.1 mM Ca^{2+} and 2.6 mM Mg^{2+}). With these divalent concentrations, the release probability is extremely low and the ΔF distribution results almost entirely from the variability in the background signal. Figure 2a plots 174 ΔF values from eight synapses and fits them with a gaussian distribution centred on zero (standard deviation (σ) = 2.8; P = 0.87, χ^2 test, degrees of freedom (d.f.) = 7). Furthermore, the distribution of ΔF values computed from trace segments before the stimulus was, as expected, similar to Fig. 2a in which stimulation leads to no release (n = 120 (7 synapses); σ = 2.7; P = 0.87, χ^2 test, d.f. = 7).

As hippocampal terminals generally release infrequently ($P_r \approx 0.3$)^{19,20} in response to one action potential in standard divalent concentrations (2 mM Ca^{2+} and 1.3 mM Mg^{2+}) we chose initially to use relatively high calcium and low magnesium bath concentrations (5 mM Ca^{2+} and 0.6 mM Mg^{2+}) to boost our

chances of getting single and double release events. We constructed a histogram of ΔF values (Fig. 2b) seen with stimulation under conditions that should have, in most cases, evoked the exocytosis of none, one or two vesicles. In this case, ΔF values were calculated by subtracting the average of the trace for 2.5 s (25 frames) before the stimulus from the average fluorescence intensity for 0.5 s (five frames) immediately after the stimulus. We used 120 measurements from eight synapses (two different preparations) to compile the histogram presented in Fig. 2b. Four evenly spaced peaks are apparent in the histogram: a large peak around zero, two positive peaks, and a small negative peak. The negative peak means that, for some trials, the fluorescence intensity decreased just after the action potential rather than increasing as expected for exocytosis.

Two phases of neurotransmitter release are recognized at hippocampal synapses: one phase (synchronous release) is tightly time-locked to the stimulus, and consists of at most a single exocytic event under the conditions of our experiments; the second phase (asynchronous release) occurs at a low rate that decreases roughly exponentially ($\tau \approx 100$ ms)²¹. As our measurements of ΔF were made on individual synapses and extend over 500 ms, we would expect to detect exocytic events from both phases. The first peak in the ΔF histogram should, therefore, correspond to release of a single vesicle, and the second peak to two vesicles. We can test this hypothesis by fitting the histogram in Fig. 2b with a theoretical curve that consists of a mixture of four gaussians characterized by two key parameters: the quantal size, q , and the quantal standard deviation, σ_q . The quantal size q is the change in fluorescence produced by the exocytosis (or endocytosis) of a single vesicle; the quantal standard deviation σ_q describes the variability of q , both within and across synapses. Of course, the means of the gaussians should be multiples of q . The standard deviation of the peak centred around zero (no release) should be σ_0 , the same as that for the histogram in Fig. 2a and the ΔF values computed before the stimulus (data not shown, σ_0 = 2.7). The standard deviation of the n th peak should be $\sqrt{\sigma_0^2 + |n|\sigma_q^2}$.

The superimposed smooth curve in Fig. 2b is the best fit, on the basis of observations combined from eight synapses, for the two free parameters q = 10.9 and σ_q = 1.76; with these parameters, the predicted (smooth curve) and observed histograms are not significantly different (P = 0.13, χ^2 test, d.f. = 11).

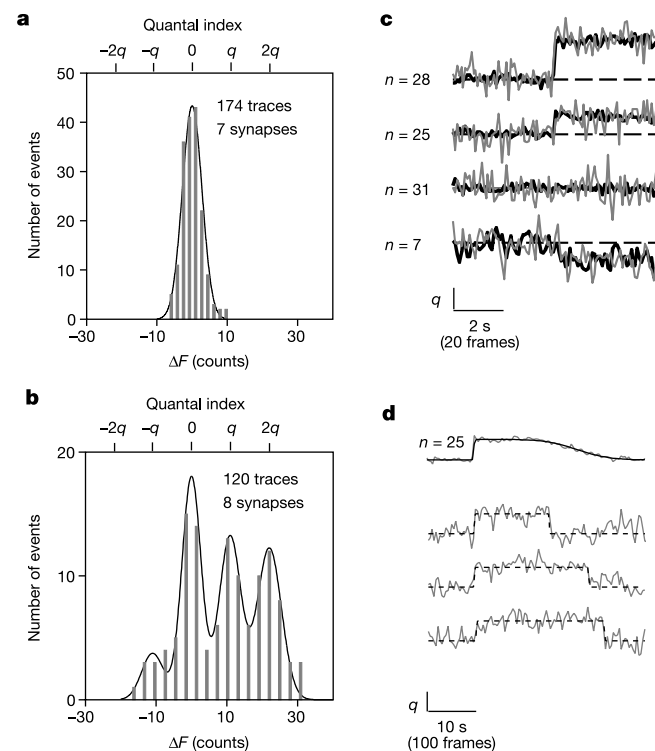


Figure 2 Detection of single-vesicle release. **a**, Histogram of signal changes, ΔF , evoked by one action potential in low calcium bath solution characterizes baseline variability. ΔF is fitted with normal distribution in black (σ = 2.8; P = 0.87, χ^2 test, d.f. = 7). **b**, ΔF histogram in high calcium conditions has four evenly spaced peaks. Overlay is the best-fitted model for quantal amplitudes constrained by quantal size q = 10.9 and quantal variability σ_q = 1.76 (P = 0.13, χ^2 test, d.f. = 11). Labels above each histogram in **a** and **b** assign ΔF to a quantal index ($-q$, 0 , $+q$, $+2q$). **c**, Averages (in black) of traces drawn from $\pm 1\sigma_0$ (σ_0 = 2.7) from each peak in **b**. Superimposed individual traces (in grey) drawn from each peak are typical of averages. **d**, Expanded time course of $+q$ average in **c** and three examples. Individual traces (signal averaged every 300 ms) return to baseline (dashed lines) more rapidly than the average. Hence, we fitted the average as a normally distributed lifetime of steps ($\mu(\tau)$ = 21.1 s; $\sigma(\tau)$ = 5.1 s).

Time course of vesicle retrieval

The preceding analysis only depends on the average change in fluorescence intensity for each release, but the time course of the fluorescence change is of paramount interest. To determine this time course, we averaged fluorescence as a function of time for the individual traces that contributed to each of the four peaks in Fig. 2b, and present these averages in Fig. 2c (black traces). A single exocytic event appears as a rapid step up with q fluorescence units as the average size. As would be expected, the average amplitude of the responses that correspond to two releases is twice as large as for one release, but the time courses are the same. To show that the averages are typical of individual events, we overlay representative individual traces (in grey) on the averages in Fig. 2c.

How long does vesicular retrieval and acidification take? The extended time course for $+q$ events in Fig. 2d shows that on average (Fig. 2d, top trace) fluorescence persists for 12.1 s before declining. However, individual traces (Fig. 2d, lower traces) persist for less or more than 12 s and then rapidly return to baseline. Therefore, the slow decay of the average time course predominantly reflects a normally distributed random wait time for vesicle retrieval.

An important test of the idea that the peaks in Fig. 2b correspond to single-vesicle cycling events is to determine whether the quantal size q is independent of release probability. We test this proposition by repeating the preceding experiment in standard rather than high extracellular calcium concentrations. But before considering these

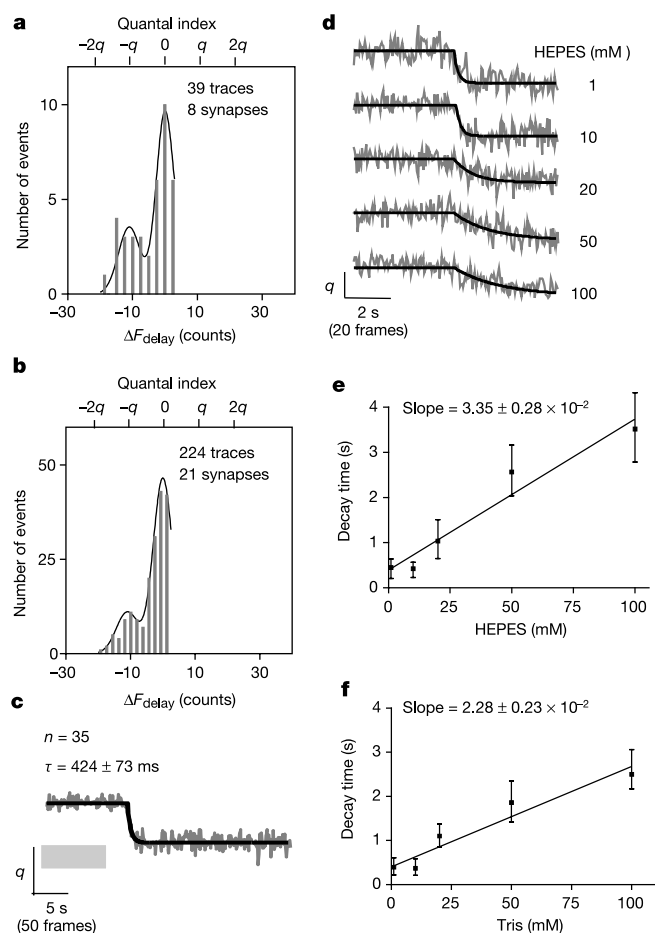


Figure 3 Evoked single-vesicle retrieval. **a**, Histogram of ΔF evoked by one action potential from Fig. 2c in a high calcium concentration bath, re-analysed to detect slower settling events. Labels above the histogram assign ΔF to quantal index obtained in Fig. 3. A peak is evident at $-q$. The histogram is clipped at $+\sigma_0$ ($\sigma_0 = 2.7$) for clarity. **b**, Tail of ΔF histogram ($\Delta F < \sigma_0$; $\sigma_0 = 2.9$) in standard calcium concentration has a peak centred at $-q$ just as for high calcium (Fig. 4a) ($P = 0.86$, χ^2 test, d.f. = 9). **c**, The average of 35 traces whose ΔF values fall within $\pm\sigma_0$ ($\sigma_0 = 2.9$) of $-q$ in **b**. The signal decays rapidly ($\tau = 424 \pm 73$ ms). **d**, Averages of ten $-q$ traces at different HEPES bath concentrations (1–100 mM) show that buffer slows $-q$ decay time. **e**, **f**, Plot of the $-q$ decay time dose responses of HEPES and Tris.

experiments, we must further examine the decreases in fluorescence seen after some stimulation trials.

Evoked single-vesicle retrieval events

The decreases in fluorescence produced by an action potential were unexpected, and we interpret them as the triggered retrieval of a single vesicle (connection with exterior severed). The basis for this interpretation is discussed below.

Although the size of the fluorescence change associated with these putative retrieval events is the same as that for single-vesicle release, the time course is different. In contrast to the rapid efflux of accumulated protons with exocytosis—as indicated by the rapid change in fluorescence associated with exocytosis (<100 ms)—the time course of these putative retrieval events would reflect both the time for retrieval and the time required for acidifying the vesicle lumen.

In the analysis described above we averaged the post-stimulus fluorescence for 0.5 s. But consideration of the time course of the average stimulus-evoked decreases in fluorescence (referred to as ‘ $-q$ events’ below) shown in Fig. 2c reveals that the transient was

still settling during the first 0.5 s. Therefore, we have re-analysed the negative tail of the histogram in Fig. 2b by averaging for a 0.5-s period starting 0.5 s after the stimulus. The re-analysed histogram (Fig. 3a) demonstrates that the $-q$ events do, indeed, have a mean amplitude (11.4 ± 0.86) that is the same as the stimulus-evoked quantal fluorescence increases.

To determine whether the $-q$ events are an artefact of the high calcium concentrations (5 mM) used above, we repeated the single action potential experiments in standard calcium concentrations (2 mM). Just as in Fig. 3a, we calculated the post-stimulus signal change by averaging the fluorescence intensity over a 0.5-s interval starting 0.5 s after the stimulus. Figure 3b presents the histogram of ΔF values elicited by an action potential. We clipped out the values greater than $+\sigma$ ($\sigma = 2.9$) from zero in the histogram because, as will be seen below, the calculation of the post-stimulus response we have used to identify $-q$ events for Fig. 3 is not appropriate for all $+q$ events in the lower calcium concentrations. As the $-q$ peak is present at standard calcium concentrations, the unitary decay signals are not an artefact of the high calcium external solution used earlier. Furthermore, the fact that the $-q$ events have the same size at two different release probabilities (and two calcium concentrations) and agree with the $+q$ amplitude supports our identification of these events as the retrieval of a single vesicle.

Time course of vesicle acidification

The time course of the $-q$ events shown in Fig. 2c must include both retrieval of the vesicle membrane from the surface and vesicular acidification. Averaging together the 35 events with ΔF values $-q \pm \sigma$ ($\sigma = 2.9$) in Fig. 3c we obtain a clearer estimate of the evoked decline in fluorescence than in Fig. 2c. The time course follows an exponential decay with a time constant of 424 ± 73 ms. The baseline is stable for a period of 10 s (100 frames) before the trace, indicating that the $-q$ events are independent of events in this period before stimulation.

If our interpretation of the $-q$ events is correct, then the addition of pH buffer to the bath solution should slow acidification of the vesicle but not materially affect the final pH value after the proton pump saturates the added buffers. We tested this idea by repeating the single action potential experiments in standard calcium concentrations while varying the HEPES concentration of the bathing solution from 1 to 100 mM. In each case the response was calculated by averaging for a 0.5-s period 8 s after the stimulus. The 8-s delay was necessary to allow the slowest condition (at 100 mM buffer) to settle. The averages of ten $-q$ event traces, selected at random, are plotted for each buffer condition in Fig. 3d with overlaid best-fitting exponentials. The slowing of the fluorescence signal decay with increasing buffer concentration is apparent, and we thus conclude that the $-q$ events represent the retrieval and acidification of a vesicle.

The time constant of acidification should increase linearly with the buffer concentration (see Supplementary Information), and this expectation is confirmed for both HEPES and Tris as shown in Fig. 3e, f. In so far as acidification is slower than the retrieval, 430 ms (the intercept of these plots) might be required to acidify the vesicle. If the endocytic step is slow, then the time constant of acidification is shorter than 430 ms, but we cannot estimate a non-trivial lower limit.

Evoked single-vesicle recycling events

As noted above, if we are correct in identifying stimulus-evoked increases in fluorescence as single-vesicle cycling events (referred to as ‘ $+q$ events’ and ‘ $+2q$ events’ below), the unitary changes in fluorescence should be independent of release probability (and of calcium concentration). To test this prediction, we repeated the analysis done for the one-action-potential, high calcium concentration experiments (Fig. 2) in standard calcium concentrations. At these divalent concentrations the release probability will be greatly

reduced compared with the earlier experiments in high calcium. As demonstrated below, the quantal size q for exocytic events is indeed independent of release probability. Unfortunately, the time course of vesicle retrieval is more complex in standard than in the higher calcium concentration, and this complicates our description.

To identify properly $+q$ events that may decay rapidly, we modified our protocol for calculating the mean post-stimulus fluorescence change by averaging the response period for 0.3 s instead of 0.5 s (three frames instead of five) after the action potential. Re-analysing the same data presented in Fig. 2b and c, we first characterized the distribution of fluorescence changes before stimulation (baseline fluctuations). The histogram plotted in Fig. 4a shows normally distributed fluorescence changes centred on zero but with a higher variability than in Fig. 2b ($\sigma = 3.3$; $n = 383$ (21 synapses); $P = 0.47$, χ^2 test, d.f. = 8). The higher variability of the baseline distribution is the expected consequence of the briefer response averaging (0.3 compared with 0.5 s). The histogram in Fig. 4b plots the distribution of fluorescence changes

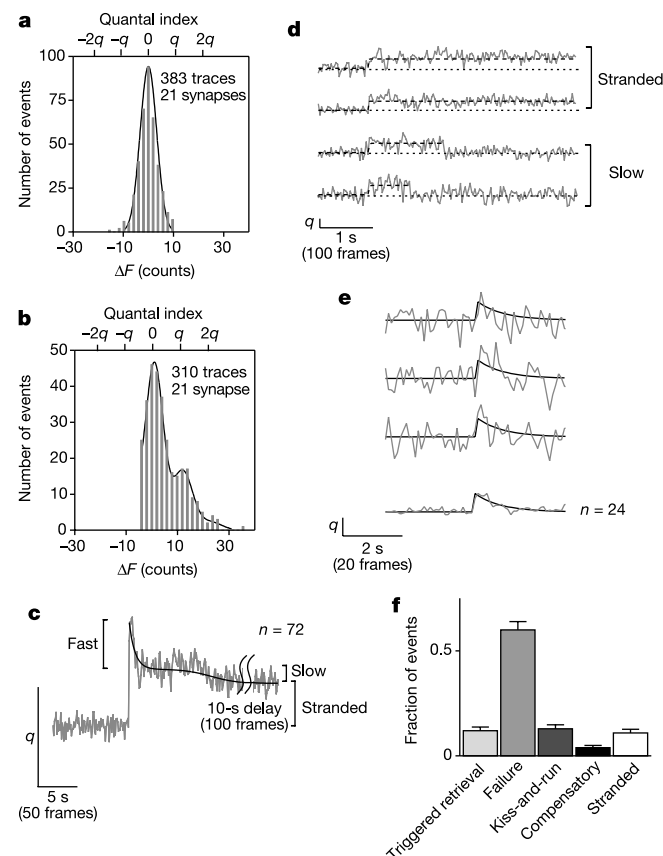


Figure 4 Kiss-and-run, compensatory and stranded events. **a**, ΔF histogram calculated without stimulation in standard calcium concentration characterizes baseline variability and is normally distributed ($\sigma = 3.3$; $P = 0.47$, χ^2 test, d.f. = 8). **b**, Using the same traces as in **a**, a ΔF histogram computed for one action potential in standard calcium has a peak at $+q$, the quantal size in high calcium concentration (see Fig. 3c) ($P = 0.99$, χ^2 test, d.f. = 16). The histogram is clipped at $-\sigma$ ($\sigma = 3.3$) for clarity. **c**, The average of 72 traces with ΔF values within $+q \pm \sigma$ (11 ± 3.4) in grey. The fit (in black) has three decay components: (1) fast, kiss-and-run monoexponential process ($\tau = 860 \pm 210$ ms); (2) slow, compensatory gaussian process ($\tau = 10.7 \pm 0.35$ s); and (3) stranded $\tau > 45$ s. The standard deviation of the compensatory fitted component constrained by the coefficient of variation found for events in high calcium (see Fig. 3d; $\sigma/\mu = 4.2$). **d**, Individual examples of compensatory and stranded recycling events (in grey). **e**, Three individual traces of kiss-and-run recycling with an average of 24 in grey; an 860-ms decay is overlaid in black. **f**, The relative frequencies of three kinds of release events, failures and evoked retrievals, in standard calcium concentration.

elicited by the stimulus in standard calcium. The overlaid smooth curve fits the distribution as a mixture of three gaussians, one peak centred at zero, another at $+q$ and a very small one at $+2q$ ($n = 310$ (21 synapses); $P = 0.99$, χ^2 test, d.f. = 16). Despite the larger standard deviation of the peak centred on zero, owing to the averaging protocol, the histogram reveals a clear exocytic peak at the same quantal size, $q = 11.0$, as in the high-calcium experiments (see Fig. 2b). We conclude that the single-vesicle fluorescence change is independent of release probability and of the extracellular calcium concentration.

Having demonstrated that q does not vary with a change in release probability, we turned our attention to the time course of the retrieval. By averaging the 72 traces falling within the ΔF interval $q \pm \sigma$ (11 ± 3.3), we obtain an average time course with three prominent decay components (Fig. 4c). The overlying best fit in black yields three time constants: a fast monoexponential process ($\tau = 860 \pm 210$ ms); a slow gaussian distributed process that persists for a minimum time (mean $\tau = 10.7 \pm 0.35$ s); and a very slow process which, at this timescale, is effectively persistent ($\tau > 45$ s).

The simplest interpretation of the decay components in the average $+q$ trace (Fig. 4c) is that there are three modes of retrieval after exocytosis, each with its own characteristic time. If a release occurs, the relative frequency of the various processes is 44% for the fast process, 20% for the slow process and 36% for failure of retrieval within 45 s. As these three putative events return to baseline at very different times it is possible to sort traces containing the three events by calculating, at appropriate delays, which traces have fallen back to within $\pm \sigma$ of zero. To isolate fast events from the $+q$ peak in Fig. 4b, for example, we collected all traces that had returned to zero ($\pm \sigma$) after a 5-s delay (value estimated from a 0.5-s average), and separated these traces from the ones that had not returned to the baseline. The slower events were separated by collecting traces that had not returned to baseline by 5 s but had by 15 s. Examination of sorted single traces reveals that three distinct modes of cycling can be seen in individual traces—Fig. 4d and e shows individual $+q$ traces that reflect the three component kinetics of the average $+q$ trace. Some $+q$ events return to baseline rapidly after stimulation (Fig. 4e), and we identify these fast events, for reasons described below, as ‘kiss-and-run’; sometimes the vesicles are retrieved slowly (Fig. 4d, bottom panel), and we call these ‘compensatory’ retrieval events (following an earlier terminology²²) because they balance membrane insertion and removal; and some exocytic events fail to terminate within the timescale of the trace (Fig. 4d, top panel), and we term these ‘stranded’ vesicles.

The bar plot in Fig. 4f summarizes the frequencies of the three types of $+q$ events relative to the failure frequency. It also includes the estimate of the evoked retrieval frequency obtained above (see Fig. 3). The evoked retrievals and stranded vesicle frequencies appear to balance each other (9% compared with 10% of all events), indicating that the two events together constitute a triggered mode of vesicle recycling.

Access to fast-cycling vesicle lumens is selective

If we correctly interpret the decay of $+q$ events as vesicular retrieval and acidification, then the time course of these decays should slow with the addition of pH buffers just as for triggered retrieval events (Fig. 5). It was not possible to test this idea for the compensatory retrieval process because its characteristic time is too long compared with the maximal effect of pH buffer on acidification time (see Fig. 3e, f). Therefore, we restricted our analysis to the fast $+q$ process (called kiss-and-run events) shown in Fig. 5a (upper trace). Using the same data from the one-action-potential experiments in standard calcium concentration analysed in Fig. 3, we found that solutions containing different concentrations of Tris buffer produced the expected slowing of the fast $+q$ decay time course (Fig. 5a, b); note, however, that the magnitude of the slowing was $70 \pm 6.7\%$ of that seen with triggered retrieval (compare Figs 3f and 5b).

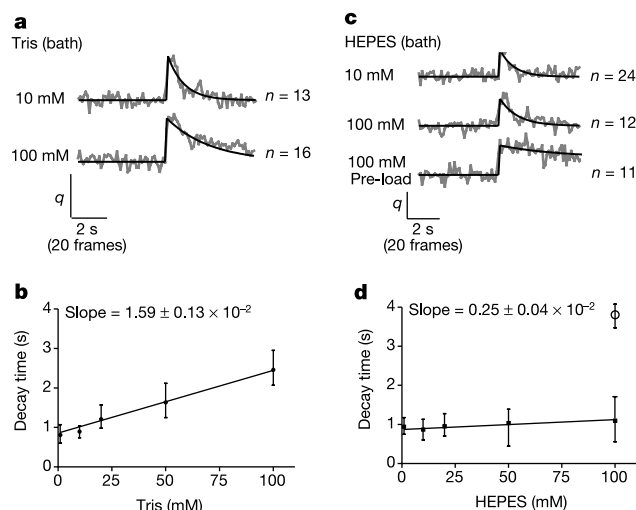


Figure 5 Reduced access of pH buffers to kiss-and-run vesicles. **a**, Average time course of isolated kiss-and-run events (in standard calcium concentration, see Fig. 4) shown for 10 mM and 100 mM Tris (in grey). Best fit monoexponential decays are superimposed in black. **b**, Plots of estimated kiss-and-run decay time across Tris concentrations. Tris slows kiss-and-run acidification, but 30% less potently than for evoked retrieval (compare slope with Fig. 3f). **c**, **d**, HEPES fails to slow kiss-and-run acidification. Averaged isolated kiss-and-run events in 10 mM and 100 mM are plotted in two upper traces in **c** with decay fits. Dose response is shown in **d** (compare with Fig. 3e). Stimulating with 900 action potentials in the presence of 100 mM HEPES before one-action-potential experiments did slow the decay time of kiss-and-run events (bottom trace in **c**).

Contrary to what was observed with triggered retrieval (and with kiss-and-run with Tris, see above), HEPES had no discernible effect on the decay of the kiss-and-run events even at concentrations up to 100 mM (Fig. 5c, d, compared with Fig. 3e). We can discount the interpretation that these events represent the superposition of evoked retrieval and compensatory recycling because we demonstrated earlier that HEPES can affect re-acidification of evoked retrieval, yet it cannot do so in these kiss-and-run events.

As Tris is effective in slowing the re-acidification of the kiss-and-run events, either HEPES is excluded from the vesicle interior by the presence of a selective fusion pore, or HEPES is, for some reason, ineffective in buffering the interior during these kiss-and-run events. We set out to test the effectiveness of HEPES in slowing kiss-and-run re-acidification by stimulating extensively (900 action potentials at 2 Hz) in the presence of 100 mM HEPES (Fig. 5c, lower trace). Our rationale was that extensive vesicle cycling can load HEPES and would slowly replace the entire recycling population with pre-loaded vesicles. The time course for the kiss-and-run events in the pre-loading, one-action-potential experiments slowed as would be predicted by the HEPES effect on evoked retrieval events (Fig. 3d, e); the decay time constant increased from 860 ms to 3.8 ± 0.35 s (Fig. 5d, open circle).

Taken together, these observations argue that fast-cycling vesicles communicate with the bath solution through a selective fusion pore and this property, together with the brief fluorescence increase, justifies our identification of the rapid events as kiss-and-run. Depending on the vesicular acidification rate, the fusion pore is open between 400 and 860 ms during kiss-and-run.

Recycling mode depends on release probability

In the preceding analysis, we have shown that q is the same within a synapse, across synapses, and for different release probabilities. Therefore, we can sort out release from failure and retrieval events for each synapse in response to one action potential and calculate its release probability. Plotted in Fig. 6a are the release probabilities for 60 terminals estimated with the criterion for a release being that the

fluorescence signal exceeds $2.5\sigma_0$ above baseline for at least 300 ms after the stimulus. As the quantal size q is approximately 3.5 times larger than σ_0 , this method will underestimate the true release probability by 10–15% (fraction of false negatives). The population of synaptic release probabilities are γ -distributed with a mean of 0.28 (with standard gamma distribution parameter $\lambda = 10.7$, order = 3). The distribution agrees qualitatively with the results of previous measurements of the broad heterogeneity of release probability in cultured hippocampal neurons^{19,20}. Although of no consequence for our conclusions, we note that the number of repeated trials per synapse (15–20) means that these experiments miss very low release probability terminals.

How frequently are the different retrieval modes used as a function of release probability? In Fig. 6b are the averaged $+q$ traces for synapses with low (mean $P_r = 0.2$), medium (mean $P_r = 0.3$) and high release probability (mean $P_r = 0.42$). Figure 6c quantifies the frequencies of the three recycling processes for each group. Low release probability synapses use kiss-and-run about 75% of the time, whereas the remainder of release events are stranded vesicles. In contrast, high release probability synapses predominantly use the compensatory mode with no stranded vesicles. The relative frequencies of the three modes in medium release probability synapses suggest that the switch from fast to slow cycling is continuous. Although the use of kiss-and-run and compensatory modes depends on release probability, the frequency of triggered retrievals is independent of release probability (Fig. 6d; $r = 0.22$). The time constants of the kiss-and-run and compensatory modes do not depend on the release probability (see Supplementary Table 1).

Discussion

By examining the exocytosis and retrieval of individual vesicles, this study helps to clarify how synaptic vesicles cycle. We find three modes of recycling and show that the release probability of a terminal determines which of these modes are used.

It was surprising that despite the pooling of data across synapses, calcium concentrations and preparations, the fluorescence signal associated with a single vesicle, q , is essentially constant. As σ_q presumably reflects the variability in the number of functional sensor proteins in each vesicle, such low variability means that the per-vesicle number of overexpressed vesicle-associated membrane protein 2 (VAMP-2) proteins, the molecule labelled with the sensor, is somehow tightly controlled. In some cases, vesicles are left on the cell surface for at least 45 s (stranded vesicle mode) and yet quantal size for retrieval remains unchanged. Apparently, then, under the conditions of our experiments, the vesicle's VAMP-2 molecules are retrieved as a unit. In secretory granules it has been demonstrated that two other membrane proteins also remain aggregated during recycling²³.

The time constants for the compensatory mode (see Supplementary Table 1) we found fits well with various reports of a 10–30 s recycling time^{4,7,9,11,12,14,17,24}; however, the finding of a minimum dwell time at the surface is new. This minimum time may reflect the time it takes for a vesicle to travel from its release to retrieval site, or it might represent the time required to assemble the molecular complex needed for retrieval (or both).

Other studies have reported an uncoupling of exocytosis and endocytosis in bulk, as we have found in stranded vesicle recycling on the unitary level^{14,17,22,25,26}. The finding that this recycling mode does not depend on probability of release but may require calcium is consistent with other findings of calcium-triggered endocytosis in the absence of recent exocytosis^{27,28}.

We interpret the time course of the $-q$ events to correspond to the time it takes to both retrieve and re-acidify a stranded vesicle. It is possible that the vesicles represented by the $-q$ events have already been retrieved by the time of the stimulus, but require a nerve impulse (presumably a calcium signal) to start acidifying. We consider this alternative interpretation to be unlikely because the

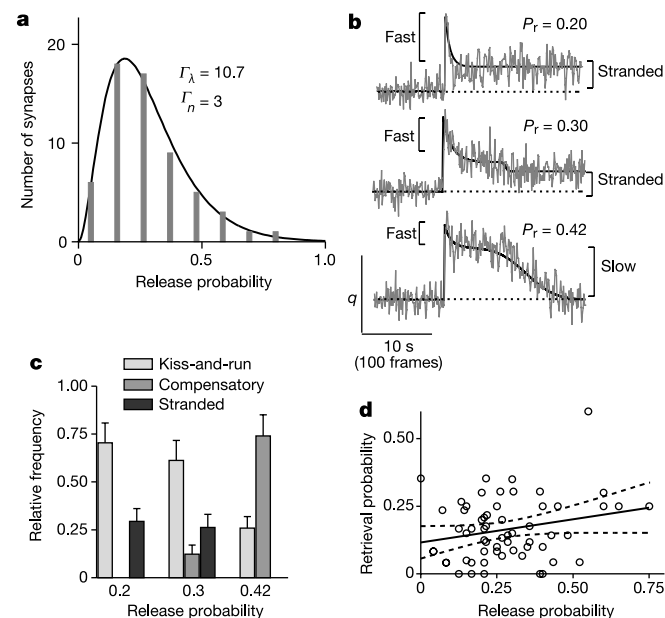


Figure 6 Balance of recycling modes depends on release probability. **a**, γ -distribution of release probability (P_r) for 60 synapses ($\mu = 0.28$; with standard gamma function parameters: $\Gamma_\lambda = 10.7$; $\Gamma_n = 3$). **b**, Average of evoked recycling traces drawn from low (mean $P_r = 0.2$; $n = 67$ traces, 20 synapses), average (mean $P_r = 0.3$; $n = 63$, 15 synapses) and high P_r synapses (mean $P_r = 0.42$; $n = 63$, 10 synapses). **c**, Quantification of the frequency of kiss-and-run, compensatory and stranded events in **b**. Low P_r terminals use the kiss-and-run mode 75% of the time or leave vesicles stranded. High P_r terminals predominantly use compensatory cycling. Average P_r synapses use all modes. **d**, Stranded retrieval probability does not depend on release probability ($r = 0.22$).

proton pump must operate constitutively to maintain the neurotransmitter gradient in the absence of increased cytosolic calcium concentrations. Nonetheless the upper bound of 430 ms is the lowest yet measured for vesicle acidification¹¹.

The kinetics of our kiss-and-run mode roughly agrees with what has been found in large terminals using capacitance-recording techniques. We estimate that it takes between 400 and 860 ms to retrieve these vesicles depending on the actual time required for acidification. In the retinal bipolar terminal, a fast recycling process takes between 300 ms and 1 s (refs 12, 17), but in the calyx of Held¹⁵ vesicles have been reported to recycle within 50 ms.

A hallmark of kiss-and-run, as first defined in secretory granule recycling, is a low conductance (<200 pS) pore open for the duration of granule fusion to the cell surface²⁹. The secretion of microvesicles, similar in size to synaptic vesicles, also reveals, albeit rarely, fusion pore conductance transients as small as 20 pS (ref. 30). In this study, we conclude that the differential entry of buffer into the vesicle lumen (Tris enters and HEPES does not) reflects the presence of a selective fusion pore. It is unclear what property of HEPES—its charge, size or shape—makes it impermeant.

Both capacitance recordings and high resolution imaging of vesicle dye release in a large synapse have failed to turn up evidence for a low permeability, kiss-and-run pore^{15,16}. However, in small hippocampal synapses, vesicles identified as using the kiss-and-run mode cannot take up or release the same lipophilic dye whereas slow cycling events can¹⁰. Perhaps large terminals behave like the high release probability synapses in this study and rarely use the kiss-and-run mode.

We find a clear shift in the balance between recycling modes as a function of release probability (Fig. 6; see also Supplementary Table 1). The lack of kiss-and-run cycling in high calcium concentrations (Fig. 2d) suggests that calcium levels are responsible for

switching a vesicle from kiss-and-run to compensatory cycling. Several reports of the effects of calcium on endocytic mode find that calcium promotes kiss-and-run^{8,12,31,32}, whereas others find no effect¹⁰ or calcium inhibition²⁶. It may be that at higher bath calcium and sustained stimulation very low release probability synapses get recruited and the cumulative amount of fast cycling appears to increase. Resolving the effect of calcium on endocytic mode and the role of release probability will be key problems for future research. □

Methods

Cell culture preparation and sensor expression

Dissociated hippocampal pyramidal cell culture was prepared from postnatal day 0 to 1 rats as described elsewhere³³. Neurons grew on an astrocyte feeder layer for 14–21 days before use. A Sindbis virus (Invitrogen) encoding the vesicle pH sensor protein was used to infect neurons. The sensor consisted of a fusion of the pH-sensitive green fluorescent protein (GFP) 'super-ecliptic phluorin' (a gift from D. DeAngelis) to the luminal aspect of rat VAMP-2 (GI: 6981614). At 16–20 h after infection, cell processes contained dimly fluorescent puncta (see Fig. 1).

Imaging set-up

Coverslips with infected cells were loaded into a field stimulation chamber (Warner; RC-21BRFS) and mounted on the stage of an Olympus IX-70 inverted microscope. Cells were illuminated with a 100 W Xe arc lamp (Optiquip) with 80% neutral density filter attenuation. Fluorescence was imaged with a $\times 60$ PlanApo 1.4 NA oil immersion objective (Olympus) onto a back-illuminated, 16-bit cooled charge-coupled device array (Roper Scientific; 512BFT) through a GFP band-pass filter set (Chroma; 41020). Images were focused onto a 128×128 element sub-region of the array. The detector electronics were configured to use 3×3 element on-chip binning, slow A-to-D readout and frame transfer mode to minimize instrument noise. The frame acquisition rate for all data is 10 Hz.

Experimental procedures

The bathing medium used in the 10 mM buffer conditions contained: 136 mM NaCl, 2.5 mM KCl, 10 mM D-glucose, 50 μ M D,L-AP5 (phosphonovaleric acid), 10 μ M 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX), and variable divalent concentrations (see below). Solutions were balanced with sucrose to 315 mmol kg⁻¹ osM⁻¹ and NaOH to a pH of 7.5. The NaCl concentration was adjusted to compensate for differing buffer concentrations, bringing it at lowest to 91 mM when 100 mM buffer was used. At this Na⁺ concentration cells could spike when stimulated with field electrodes (data not shown). Three divalent cation concentrations were used in the experiments: low calcium, 0.1 mM CaCl₂ and 2 mM MgCl₂; standard calcium, 2 mM CaCl₂ and 1.3 mM MgCl₂; and high calcium, 5 mM CaCl₂ and 0.6 mM MgCl₂.

Once a suitable field of view was found, the main experiments began with a round of photobleaching, done using the same illumination intensity as for recordings. The photobleaching time constant at this intensity was approximately 14 min (8,400 frames, data not shown). Most terminals were bleached for three rounds, 14 min each.

Stimulation consisted of a 1 ms 10 V cm⁻¹ pulse across the chamber electrodes delivered using a Grass stimulator (SD9) synchronized by computer control with the image acquisition. The presence of CNQX and AP5 blocked recurrent action potentials. Each shock generated only one nerve impulse, as confirmed by cell-attached patch recordings (data not shown).

After photobleaching was complete the main recording session could begin. Each session consisted of 3–4 time periods, 4–5 repeats each, of 1 min recordings (600 frames). One shock was delivered time-locked to the 150th frame. The period between repeats was approximately 30 s. For every time period, a 20-action-potential stimulation (at 20 Hz) was delivered starting at frame 150 to confirm the stability of the responses. Any time periods in which the 20-action-potential response failed to match earlier events within 66% were eliminated from further analysis.

Image analysis

Data were acquired using WinView software (Roper Scientific), analysed using custom-written Matlab routines (Mathworks) and plotted in Prism (Graphpad). To find active synapses, we subtracted frame 149 from 160, the frames before and after the 20-action-potential stimulation, respectively. The difference image showed spots that roughly corresponded to location of background fluorescence in the unsubtracted images (see Fig. 1). 4×4 pixel-sized (12×12 elements; $2.68 \times 2.68 \mu$ m) ROIs were centred on difference image puncta using a centroid centring algorithm. We chose the ROI size empirically to accommodate spatial drift during repeated trials and to maximize the integrated signal change evoked by the 20-action-potential probe while minimizing background noise. The ROIs determined from the difference image were moved to the frames of subsequent experiments. The pixel values in the ROI were summed for each frame and assembled into a 10-Hz trace. The residual photobleaching was de-trended from the trace, calculated by fitting the 15-s (150 frame) background segment beginning each trace with a 15-min (8,400 frame) exponential decay (data not shown).

All error bars report standard error of the mean (s.e.m.). χ^2 statistics for the difference between observed and predicted histograms of fluorescence change were computed using only the bins containing five or more events.

Received 16 January; accepted 4 April 2003; doi:10.1038/nature01677.

1. Ryan, T. A. & Smith, S. J. Vesicle pool mobilization during action potential firing at hippocampal synapses. *Neuron* **14**, 983–989 (1995).
2. Harata, N. *et al.* Limited numbers of recycling vesicles in small CNS nerve terminals: implications for neural signaling and vesicular cycling. *Trends Neurosci.* **24**, 637–643 (2001).
3. Ryan, T. A. *et al.* The kinetics of synaptic vesicle recycling measured at single presynaptic boutons. *Neuron* **11**, 713–724 (1993).
4. von Gersdorff, H. & Matthews, G. Dynamics of synaptic vesicle fusion and membrane retrieval in synaptic terminals. *Nature* **367**, 735–739 (1994).
5. von Gersdorff, H. & Matthews, G. Inhibition of endocytosis by elevated internal calcium in a synaptic terminal. *Nature* **370**, 652–655 (1994).
6. Wu, L. G. & Betz, W. J. Nerve activity but not intracellular calcium determines the time course of endocytosis at the frog neuromuscular junction. *Neuron* **17**, 769–779 (1996).
7. Ryan, T. A., Smith, S. J. & Reuter, H. The timing of synaptic vesicle endocytosis. *Proc. Natl Acad. Sci. USA* **93**, 5567–5571 (1996).
8. Klingauf, J., Kavalali, E. T. & Tsien, R. W. Kinetics and regulation of fast endocytosis at hippocampal synapses. *Nature* **394**, 581–585 (1998).
9. Pyle, J. L., Kavalali, E. T., Piedras-Renteria, E. S. & Tsien, R. W. Rapid reuse of readily releasable pool vesicles at hippocampal synapses. *Neuron* **28**, 221–231 (2000).
10. Stevens, C. F. & Williams, J. H. 'Kiss and run' exocytosis at hippocampal synapses. *Proc. Natl Acad. Sci. USA* **97**, 12828–12833 (2000).
11. Sankaranarayanan, S. & Ryan, T. A. Real-time measurements of vesicle-SNARE recycling in synapses of the central nervous system. *Nature Cell Biol.* **2**, 197–204 (2000).
12. Neves, G., Gomis, A. & Lagnado, L. Calcium influx selects the fast mode of endocytosis in the synaptic terminal of retinal bipolar cells. *Proc. Natl Acad. Sci. USA* **98**, 15282–15287 (2001).
13. Sankaranarayanan, S. & Ryan, T. A. Calcium accelerates endocytosis of vSNAREs at hippocampal synapses. *Nature Neurosci.* **4**, 129–136 (2001).
14. Heidelberger, R. ATP is required at an early step in compensatory endocytosis in synaptic terminals. *J. Neurosci.* **21**, 6467–6474 (2001).
15. Sun, J. Y., Wu, X. S. & Wu, L. G. Single and multiple vesicle fusion induce different rates of endocytosis at a central synapse. *Nature* **417**, 555–559 (2002).
16. Zenisek, D., Steyer, J. A., Feldman, M. E. & Almers, W. A membrane marker leaves synaptic vesicles in milliseconds after exocytosis in retinal bipolar cells. *Neuron* **35**, 1085–1097 (2002).
17. Heidelberger, R., Zhou, Z. Y. & Matthews, G. Multiple components of membrane retrieval in synaptic terminals revealed by changes in hydrostatic pressure. *J. Neurophysiol.* **88**, 2509–2517 (2002).
18. Miesenböck, G., De Angelis, D. A. & Rothman, J. E. Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins. *Nature* **394**, 192–195 (1998).
19. Murthy, V. N., Sejnowski, T. J. & Stevens, C. F. Heterogeneous release properties of visualized individual hippocampal synapses. *Neuron* **18**, 599–612 (1997).
20. Huang, E. P. & Stevens, C. F. Estimating the distribution of synaptic reliabilities. *J. Neurophysiol.* **78**, 2870–2880 (1997).
21. Goda, Y. & Stevens, C. F. Two components of transmitter release at a central synapse. *Proc. Natl Acad. Sci. USA* **91**, 12942–12946 (1994).
22. Smith, C. & Neher, E. Multiple forms of endocytosis in bovine adrenal chromaffin cells. *J. Cell Biol.* **139**, 885–894 (1997).
23. Taraska, J. W., Perrais, D., Ohara-Imaizumi, M., Nagamatsu, S. & Almers, W. Secretory granules are

recaptured largely intact after stimulated exocytosis in cultured endocrine cells. *Proc. Natl Acad. Sci. USA* **100**, 2070–2075 (2003).

24. Beutner, D., Voets, T., Neher, E. & Moser, T. Calcium dependence of exocytosis and endocytosis at the cochlear inner hair cell afferent synapse. *Neuron* **29**, 681–690 (2001).
25. Engisch, K. L. & Nowycky, M. C. Compensatory and excess retrieval: two types of endocytosis following single step depolarizations in bovine adrenal chromaffin cells. *J. Physiol.* **506**, 591–608 (1998).
26. Hsu, S. F. & Jackson, M. B. Rapid exocytosis and endocytosis in nerve terminals of the rat posterior pituitary. *J. Physiol.* **494**, 539–553 (1996).
27. Neale, E. A., Bowers, L. M., Jia, M., Bateman, K. E. & Williamson, L. C. Botulinum neurotoxin A blocks synaptic vesicle exocytosis but not endocytosis at the nerve terminal. *J. Cell Biol.* **147**, 1249–1260 (1999).
28. Thomas, P., Lee, A. K., Wong, J. G. & Almers, W. A triggered mechanism retrieves membrane in seconds after Ca^{2+} -stimulated exocytosis in single pituitary cells. *J. Cell Biol.* **124**, 667–675 (1994).
29. Albillos, A. *et al.* The exocytotic event in chromaffin cells revealed by patch amperometry. *Nature* **389**, 509–512 (1997).
30. Klyachko, V. A. & Jackson, M. B. Capacitance steps and fusion pores of small and large-dense-core vesicles in nerve terminals. *Nature* **418**, 89–92 (2002).
31. Ales, E. *et al.* High calcium concentrations shift the mode of exocytosis to the kiss- and-run mechanism. *Nature Cell Biol.* **1**, 40–44 (1999).
32. Cousin, M. A. & Robinson, P. J. Two mechanisms of synaptic vesicle recycling in rat brain nerve terminals. *J. Neurochem.* **75**, 1645–1653 (2000).
33. Murthy, V. N. & Stevens, C. F. Reversal of synaptic vesicle docking at central synapses. *Nature Neurosci.* **2**, 503–507 (1999).
34. Sankaranarayanan, S., De Angelis, D., Rothman, J. E. & Ryan, T. A. The use of pHluorins for optical measurements of presynaptic activity. *Biophys. J.* **79**, 2199–2208 (2000).
35. Wu, M. M. *et al.* Mechanisms of pH regulation in the regulated secretory pathway. *J. Biol. Chem.* **276**, 33027–33035 (2001).
36. Wu, M. M. *et al.* Organelle pH studies using targeted avidin and fluorescein-biotin. *Chem. Biol.* **7**, 197–209 (2000).
37. Nishi, T. & Forgac, M. The vacuolar (H^{+})-ATPases—nature's most versatile proton pumps. *Nature Rev. Mol. Cell Biol.* **3**, 94–103 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank R. Jacobs, M. Pilla and J. Sullivan for technical support; D. De Angelis for the super-ecliptic phluorin construct; Roger Tsien for suggesting the use of photobleaching; and N. Spitzer for a critical reading of this manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.F.S. (cfs@salk.edu).

Insights into IgA-mediated immune responses from the crystal structures of human Fc α RI and its complex with IgA1-Fc

Andrew B. Herr*, Edward R. Ballister*†‡ & Pamela J. Bjorkman*§

* Division of Biology 114-96 and § Howard Hughes Medical Institute, California Institute of Technology, Pasadena, California 91125, USA

† La Salle High School, Pasadena, California 91107, USA

‡ Present address: University of California at Berkeley, Berkeley, California 94720, USA

Immunoglobulin- α (IgA)-bound antigens induce immune effector responses by activating the IgA-specific receptor Fc α RI (CD89) on immune cells. Here we present crystal structures of human Fc α RI alone and in a complex with the Fc region of IgA1 (Fc α). Fc α RI has two immunoglobulin-like domains that are oriented at approximately right angles to each other. Fc α resembles the Fcs of immunoglobulins IgG and IgE, but has differently located interchain disulphide bonds and external rather than interdomain *N*-linked carbohydrates. Unlike 1:1 Fc γ RIII:IgG and Fc ϵ RI:IgE complexes, two Fc α RI molecules bind each Fc α dimer, one at each C α 2–C α 3 junction. The Fc α RI-binding site on IgA1 overlaps the reported polymeric immunoglobulin receptor (pIgR)-binding site, which might explain why secretory IgA cannot initiate phagocytosis or bind to Fc α RI-expressing cells in the absence of an integrin co-receptor.

Adult humans produce more IgA per day than all other antibody isotypes combined¹. IgA is the predominant antibody isotype in mucosal regions, which represent the primary avenue for invasion by many pathogens. The mucosal immune system is responsible for multiple levels of protection against pathogens: induction of immunity in mucosa-associated lymphoid tissues, prevention of pathogen adherence by secretory IgA-mediated immune exclusion, and systemic immune effector functions through the action of serum antibodies².

Two distinct forms of IgA are found in serum: monomeric and dimeric IgA. Monomeric IgA is analogous to IgG and IgE, but has additional carboxy-terminal tailpieces that interact with J- (joining) chain, forming covalent dimeric IgA³. Co-expression of IgA and J-chain by B cells in the mucosal lamina propria leads to high local concentrations of dimeric IgA in these regions, whereas serum IgA is predominantly monomeric¹. Dimeric IgA is specifically bound by pIgR on the basolateral surfaces of the mucosal epithelium, transcytosed through the epithelial layer, and released into mucosal secretions as a covalent complex of dimeric IgA and the cleaved pIgR ectodomain (called bound secretory component, SC)³. The complex of dimeric IgA with bound SC is known as secretory IgA (SIgA) and represents a third distinct form of IgA.

IgA-mediated immune effector responses such as phagocytosis, antibody-dependent cell-mediated cytotoxicity, respiratory burst and cytokine release are mediated through Fc α RI (CD89), an IgA-specific receptor that is expressed on monocytes, eosinophils, neutrophils and macrophages⁴. Both monomeric and dimeric IgA can bind to Fc α RI, and monomeric or dimeric IgA immune complexes can activate phagocytosis and other immune responses through the clustering of Fc α RI⁵. SIgA cannot trigger phagocytosis^{5,6}, although it can initiate certain inflammatory responses, such as respiratory burst in polymorphonuclear leukocytes⁷ and degranulation of eosinophils⁸, through Fc α RI and an integrin co-receptor.

Although Fc α RI is distantly related to Fc γ R and Fc ϵ RI proteins, it shares higher sequence similarity with the leukocyte immunoglobulin-like receptors (LIR/LILR/ILTs), killer inhibitory receptors

(KIRs) and other members of the leukocyte receptor cluster⁹. Unlike Fc γ RIII and Fc ϵ RI, which form 1:1 receptor:Fc complexes through interactions with hinge-proximal Fc regions^{10,11}, Fc α RI forms a 2:1 complex with an Fc α dimer¹² by binding to each C_H2–C_H3 domain interface (called C α 2 and C α 3 in IgA)^{13,14}, reminiscent of the 2:1 complex formed between the neonatal Fc receptor (FcRn) and IgG, in which FcRn binds to the C_H2–C_H3 interface of Fc γ ¹⁵. To understand the molecular basis of the Fc α RI:IgA interaction, we determined the crystal structures of the Fc α RI ectodomain in two crystal forms and a 2:1 Fc α RI:Fc α complex at resolutions of 3.0, 3.9 and 3.1 Å, respectively (Supplementary Table 1).

Structure of Fc α RI

The Fc α RI ectodomain is composed of two immunoglobulin-like domains oriented at approximately 90° to one another (Fig. 1a). The domains of Fc α RI have a similar folding topology to the domains of LIR-1 (ref. 16) and other immunoglobulin superfamily Fc receptors^{11,17–19} (Fig. 1b; see Supplementary Table 2 for r.m.s. deviations between the domains of Fc α RI and other receptors). In both domains 1 and 2 (D1 and D2), the first β -strand is shared between the two β -sheets. D1 includes short C' and D strands connecting strands C and E, whereas there is no D strand in D2. There are short stretches of 3₁₀ helix in both domains, and a polyproline type-II helix in D2 (Fig. 1a). Ordered carbohydrate is visible at two of the four potential *N*-linked glycosylation sites in each molecule of free Fc α RI in the 3.0 Å structure (N120 and N156 in the first molecule, and N44 and N58 in the second molecule) and at all four sites in the Fc α -bound Fc α RI structure. The overall domain orientation resembles that seen in the LIR-1 (ref. 16; Fig. 1c) and KIR²⁰ receptors. Although other immunoglobulin superfamily Fc receptors such as Fc ϵ RI, Fc γ RIIa, Fc γ RIIb and Fc γ RIII also have overall bent structures^{11,17–19}, the relative D1–D2 orientation in these receptors is opposite to that observed for Fc α RI and LIR-1 (ref. 16; Fig. 1c).

A comparison of three independent structures of free Fc α RI (two molecules from the 3.0 Å structure, and one from the 3.9 Å struc-

ture) reveals moderate flexibility in the D1–D2 interface; the interdomain angles are 97° and 92° for the two molecules in the 3.0 Å structure, and 88° for the low-resolution model. The D1–D2 interface buries 1,170–1,238 Å², slightly more than interdomain interface areas found in the KIR (1,154 Å²; ref. 20) and LIR-1 (1,012 Å²; ref. 16) receptors. The DE and FG loops in D1 show variable conformations in the independent views of free FcαRI and in FcαRI complexed with Fcα (Fig. 1d).

Structure of Fcα

The structure of Fcα described here is derived from the FcαRI:Fcα complex structure, because attempts to crystallize Fcα alone were unsuccessful. Fcα is a two-fold symmetric dimer of IgA heavy chains, each with two immunoglobulin constant domains, Cα2 and Cα3 (Fig. 2). The Cα3 domains form an extensive dimer interface that buries a total of 2,061 Å² and associate to form similar dimerization contacts to those seen in the analogous Cγ3–Cγ3 (Fig. 2a) and Cε4–Cε4 interactions in Fcγ²¹ and Fcε^{22,23}, respectively.

Apart from minor variations in loop regions, the primary differences between Fcα and other Fcs of known structure (Fcγ and Fcε) involve the positions of interdomain disulphide bonds and *N*-linked carbohydrates. Like Fcγ, the heavy chains of Fcα are covalently coupled through disulphides linking hinge residues (C241 to counterpart C241; not present in our construct)^{24,25}. However, Fcα also has unusual disulphides between C299 in the DE loop of each Cα2 domain and C242 at the base of the hinge in the opposite heavy chain (Fig. 2b). These interdomain disulphide bonds are reminiscent of those observed linking the IgE heavy

chains at the base of the Cε2 domains²², although the residues involved are not analogous. Interestingly, Fcα C299 is the counterpart of Fcγ N297 and Fcε N394, the residues to which the interdomain *N*-linked glycans that stabilize the Cγ2 or Cε3 domains are attached. Thus, this substitution in Fcα of cysteine for asparagine allows disulphide bond formation and prevents attachment of *N*-linked glycans between the Cα2 domains, which is a conserved feature of Cγ2 domains and the counterpart Cε3 domains. Instead, Fcα contains *N*-linked carbohydrates attached to N263 on an outer surface of the Cα2 domains (Fig. 2a). Each *N*-glycan contacts both the Cα2 and Cα3 domains, burying 465 Å² and 457 Å², respectively, unlike the Fcγ and Fcε *N*-glycans, which contact only the Cγ2 (or Cε3) domains.

Structure of the FcαRI:Fcα complex

The FcαRI:Fcα complex consists of two FcαRI molecules bound to a single Fcα dimer, with one receptor binding at each Cα2–Cα3 interface (Fig. 3a, b). The FcαRI-binding site on Fcα is composed of residues in the AB helix/loop of the Cα2 domain (L256, L257, L258) and several regions of the Cα3 domain: the A strand (E348), the C strand (R382, L384), the CC' loop (S387, E389), the F strand (M433, H436), the FG loop (E437, A438, L439, P440, L441) and the G strand (A442, F443, T444, Q445) (Figs 2b and 4a, and Supplementary Table 3). These residues are conserved in all human IgA subclasses and allotypes (IgA1, IgA2m(1), IgA2m(2) and IgA2n). Sequence alignment of IgA1 with other antibody isotypes shows that the specificity of FcαRI for IgA appears to arise predominantly from residues in the C strand and CC' loop,

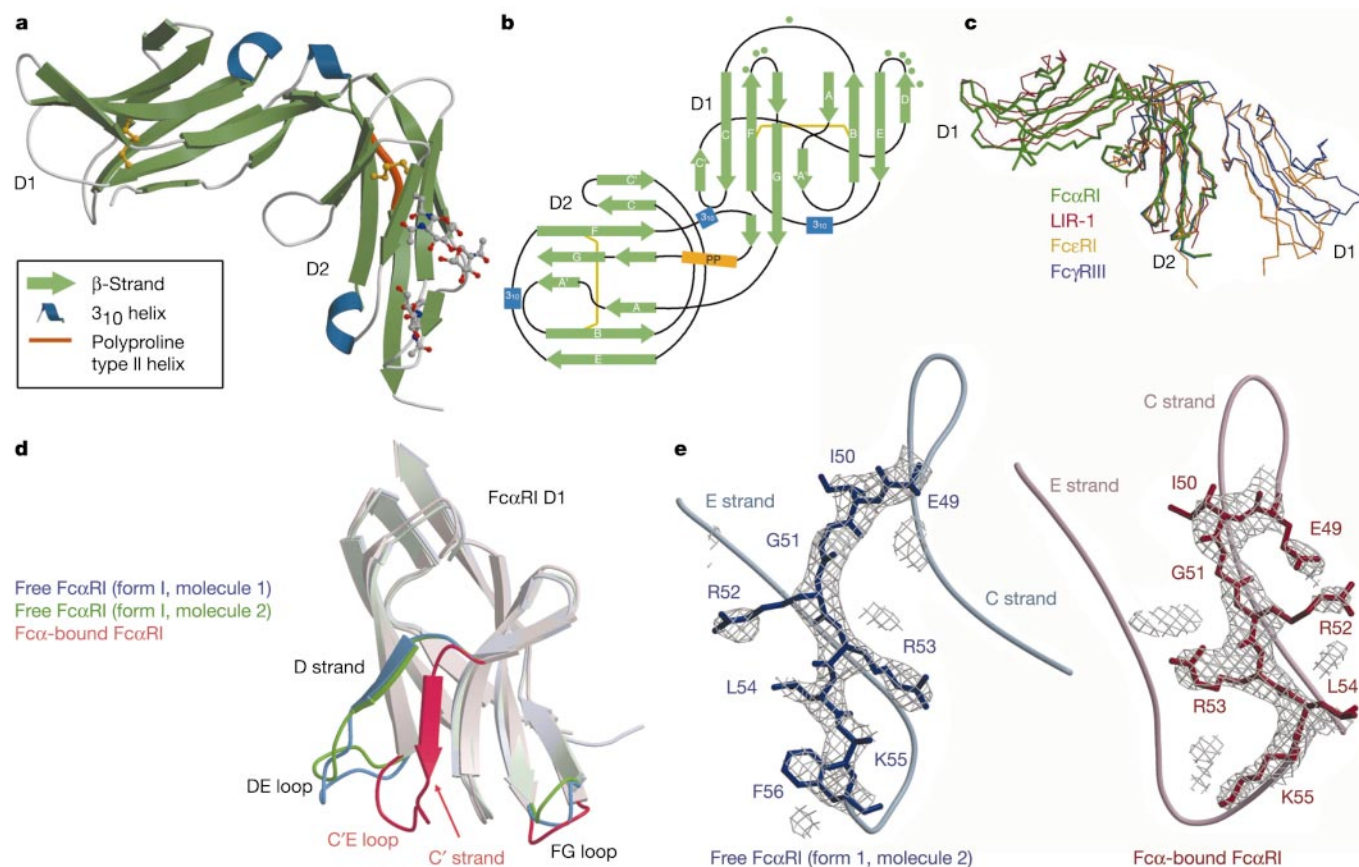


Figure 1 FcαRI structure. **a**, Ribbon diagram of FcαRI. Disulphide bonds are yellow and carbohydrate residues are shown in ball-and-stick representation. **b**, FcαRI topology diagram. β-Strands are green, 3_{10} helices are blue, polyproline type-II helices are orange and disulphide bonds are yellow. Green dots show residues that contact Fcα. **c**, Superimposed Cα traces of FcαRI, LIR-1 (ref. 16), FcεRI¹⁷ and FcγRIII¹¹.

d, Superposition of the D1 domains in independently determined FcαRI structures. Regions that differ are highlighted in blue, green or red. **e**, Simulated annealing electron density omit maps showing the D or C' strands in D1 of free and Fcα-bound FcαRI, respectively, with ball-and-stick models of the deleted regions superimposed.

with contributions from residues in the F strand, FG loop and G strand (Fig. 5). The site on Fc α RI that binds Fc α involves residues found in the BC loop (Y35), the D strand (R52, R53, L54, K55), the DE loop (F56, W57, carbohydrate linked to N58), and the FG loop (R82, G84, H85) of D1 (Figs 1b and 4a, and Supplementary Table 3). A similar region of LIR-1 has been mapped as the binding site for UL18, a viral major histocompatibility complex (MHC) class I homologue¹⁶.

The Fc α RI:Fc α interface is composed of a central hydrophobic core (Fc α RI residues Y35, L54, F56, G84, H85 and the aliphatic portion of K55 packing against Fc α residues L257, L258, M433, L441, A442, F443 and the aliphatic portion of the R382 side chain) flanked by charged residues (Fc α RI R52, R53, K55 and R82, and Fc α R382, E389 and E437) (Fig. 4b), burying a total of 1,656 Å². Several potential intermolecular hydrogen bonds are found at the periphery of the site, but no salt bridges are observed, consistent with binding studies showing that <10% of the free energy of binding is contributed by electrostatic interactions¹². The *N*-glycan attached to Fc α RI N58 forms two potential hydrogen bonds and a van der Waals contact with Fc α (Fig. 4a). Although the observed *N*-glycans on Fc α approach within 8 Å of Fc α RI, they do not directly contact the receptor.

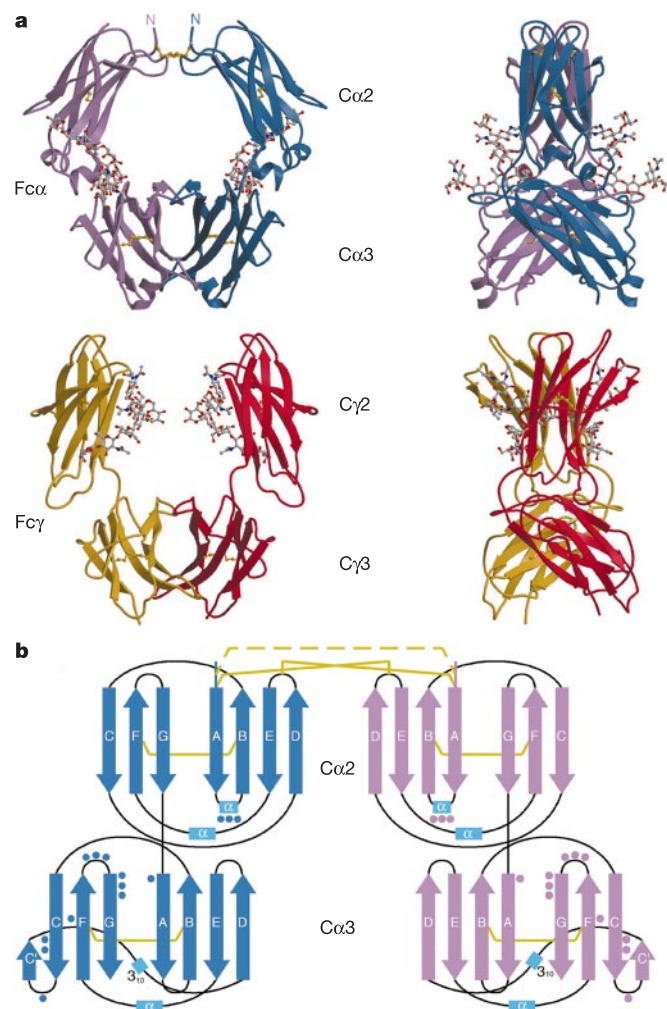


Figure 2 Fc α structure. **a**, Ribbon diagrams showing front (left) and side (right) views of Fc α (top) and Fc γ (bottom)³³. Disulphide bonds are shown in yellow and carbohydrate residues are shown in ball-and-stick representation. **b**, Topology diagram of Fc α . β -Strands are blue or magenta, 3_{10} and α -helices are light blue and disulphides are yellow. The proposed C241–C241 disulphide bond (not present in our construct) is shown as a dashed yellow line. Blue and magenta dots show residues that contact Fc α RI.

The interaction surface identified by the Fc α RI:Fc α structure can be used to interpret mutagenesis studies^{13,14,26–28} (Fig. 4c). Substitution of Fc α RI Y35, R82 or Y81 to alanine, or H85 to arginine resulted in 100-fold loss of affinity or ablation of binding^{27,28}. Y35 is located at the centre of the hydrophobic patch in the binding site and forms a potential hydrogen bond to the carbonyl oxygen of Fc α L257. R82 forms a hydrogen bond with the carbonyl oxygen of Fc α L256. Y81 is located in the hydrophobic core of Fc α RI D1 and is likely to be essential for proper folding of the domain. H85 interacts with the Fc α FG loop. Mutations of Fc α RI R52, I83 or G84 to alanine resulted in 8- to 19-fold losses of affinity²⁸. R52 forms two potential hydrogen bonds with the carbonyl oxygen of L258, G84 contacts the FG loop, and I83 appears to stabilize the conformation of the Fc α RI FG-binding loop, although it does not contact Fc α . Mutation of Fc α residues in the AB helix/loop (L257R or L258R), the FG loop (P440A, L441R, A442R, L441M/A442N) or the G strand (F443R) resulted in ablation of binding in rosetting assays^{13,14}; these residues are all involved in the observed Fc α RI:Fc α interface, and all but one contribute to the hydrophobic core of the binding site.

The Fc α RI:Fc α structure reveals other residues in the binding site that had not been identified previously, including Fc α RI residues in the D strand (L54, K55, F56, W57) and N58-linked sugar, and Fc α residues in the C α 2 AB helix/loop (L256), the C α 3 A (E348), C (R382, L384), F (M433, H436) and G (Q445) strands, and the CC' (S387, E389) and FG loops (E437, A438, L439). In particular, Fc α RI residues L54, K55 and F56 seem to be important for binding, as each makes three to five contacts with Fc α and contributes 16%, 9% or 10%, respectively, of the buried surface area in the interface (Supplementary Table 3).

Conformational changes on complex formation

Several conformational changes occur within Fc α RI D1 on binding Fc α . Superposition of D1 in the free and bound receptors shows that the Fc α RI D strand and the DE and FG loops adopt different conformations (Fig. 1d, e). The D strand in free Fc α RI shifts from the front β -sheet (strands ABED) to the back β -sheet (strands A'GFCC') on binding Fc α , forming instead a C' strand in the complex. The main secondary structural elements of Fc α RI that participate in binding to Fc α (the D/C' strand and the DE/C'E and FG loops) show inherent flexibility in free Fc α RI, and they adopt new conformations on binding (Fig. 1d). A comparison of the structure of Fc ϵ RI in several crystal forms showed similar structural variability in the analogous regions of D2 (ref. 29). No significant change in the angle between Fc α RI domains occurs on binding to Fc α ; the D1–D2 angle when bound to Fc α (96°) is nearly identical to that observed for one molecule of free Fc α RI (97°) and similar to the other two molecules of free Fc α RI (88°, 92°).

Free Fc γ ²⁴ and Fc ϵ ^{22,23} show significant conformational changes on binding to Fc γ RIII¹¹ or Fc ϵ RI¹⁰, respectively. These conformational changes induce asymmetry in the Fc molecules when the FcR binds. By contrast, the Fc α dimer in the Fc α RI:Fc α structure is two-fold symmetric. Since the structure of free Fc α is not known, it is not clear whether any conformational changes occur when Fc α RI binds. However, the disulphide bonds at the top of the C α 2 domains and the large interface between the C α 3 domains burying a total of 2,061 Å² would be expected to constrain the conformational freedom of Fc α .

Comparison to other FcR:Fc interactions

The C α 2–C α 3 interface of Fc α where Fc α RI binds is analogous to the C γ 2–C γ 3 interface on Fc γ that serves as the binding site for a number of structurally distinct proteins, including FcRn¹⁵ (Fig. 3c), protein A³⁰, protein G³¹, rheumatoid factor³² and Fc-III, a peptide selected for high-affinity IgG binding³³. The binding of these proteins to Fc γ involves residues in the AB loop of the C γ 2 domain and the G strand of the C γ 3 domain that create a hydrophobic and

highly accessible surface³³. Although the Fc α RI-binding site on Fc α also incorporates these regions, none of the residues is conserved exactly between Fc α RI:Fc α and the other complexes. However, the Fc α and Fc γ binding surfaces share the common property of being primarily hydrophobic with outlying charged residues.

Fc α RI is structurally similar to Fc γ RIII and Fc ϵ RI, yet the Fc α RI:Fc α binding interaction is completely different from the Fc γ RIII:Fc γ and Fc ϵ RI:Fc ϵ complexes, in which a single Fc receptor binds across both C γ 2 (or C ϵ 3) domains in a region adjacent to the Fc γ hinge or the Fc ϵ C ϵ 2 domains^{10,11} (Fig. 3d, e). Furthermore, Fc α is bound to Fc α RI in an 'upright' orientation with its C termini presumably near the membrane, whereas both Fc γ and Fc ϵ are bound by their receptors in the opposite orientation^{10,11} (Fig. 3). The 'upside-down' orientation of receptor-bound IgG and IgE antibodies probably does not cause steric hindrance between the Fabs and the cell surface, because the hinge region of IgG antibodies is highly flexible³⁴, and the Fc region of IgE has a sharp bend between the C ϵ 2 and C ϵ 3–C ϵ 4 domains²². By contrast, IgA1 has a heavily O-glycosylated hinge region, which probably adopts a rigid conformation³⁵. Thus, the IgA1 hinge is unlikely to be flexible enough to allow binding in the mode observed in the Fc γ RIII:Fc γ and Fc ϵ RI:Fc ϵ complexes. It is potentially relevant that IgA2 has a 13-residue deletion in the hinge region, eliminating the O-glycosy-

lated region¹; binding of Fc α RI to the C α 2–C α 3 junction therefore allows identical interactions with both subclasses of IgA, which would be unlikely if Fc α RI bound IgA in a hinge-proximal region.

Implications for signalling through Fc α RI

Both monomeric and dimeric forms of serum IgA are able to bind and initiate immune effector functions through Fc α RI⁵. The primary function of SIgA seems to be immune exclusion, an Fc α RI-independent process in which SIgA binds to pathogens and sterically prevents their adherence to the mucosal epithelia^{1,2}. SIgA is unable to trigger phagocytosis^{5,6}, although it can initiate respiratory burst in polymorphonuclear leukocytes⁷, and degranulation of basophils and eosinophils^{8,36}. Binding of polymorphonuclear leukocytes to immobilized SIgA⁷ and degranulation of eosinophils⁸ both show an absolute requirement for the integrin co-receptor Mac-1 (or its β -subunit, CD18), suggesting that SIgA is unable to bind or activate Fc α RI alone without the integrin co-receptor. The inability of SIgA to bind or activate Fc α RI can be explained based on the structural information presented here. Bound SC in SIgA interacts with IgA residues in the DE and FG loops^{37,38} and forms a disulphide bond with C311 of IgA³. As described above, the FG loop comprises a major part of the Fc α RI-binding site on Fc α (contributing 23% of the total binding

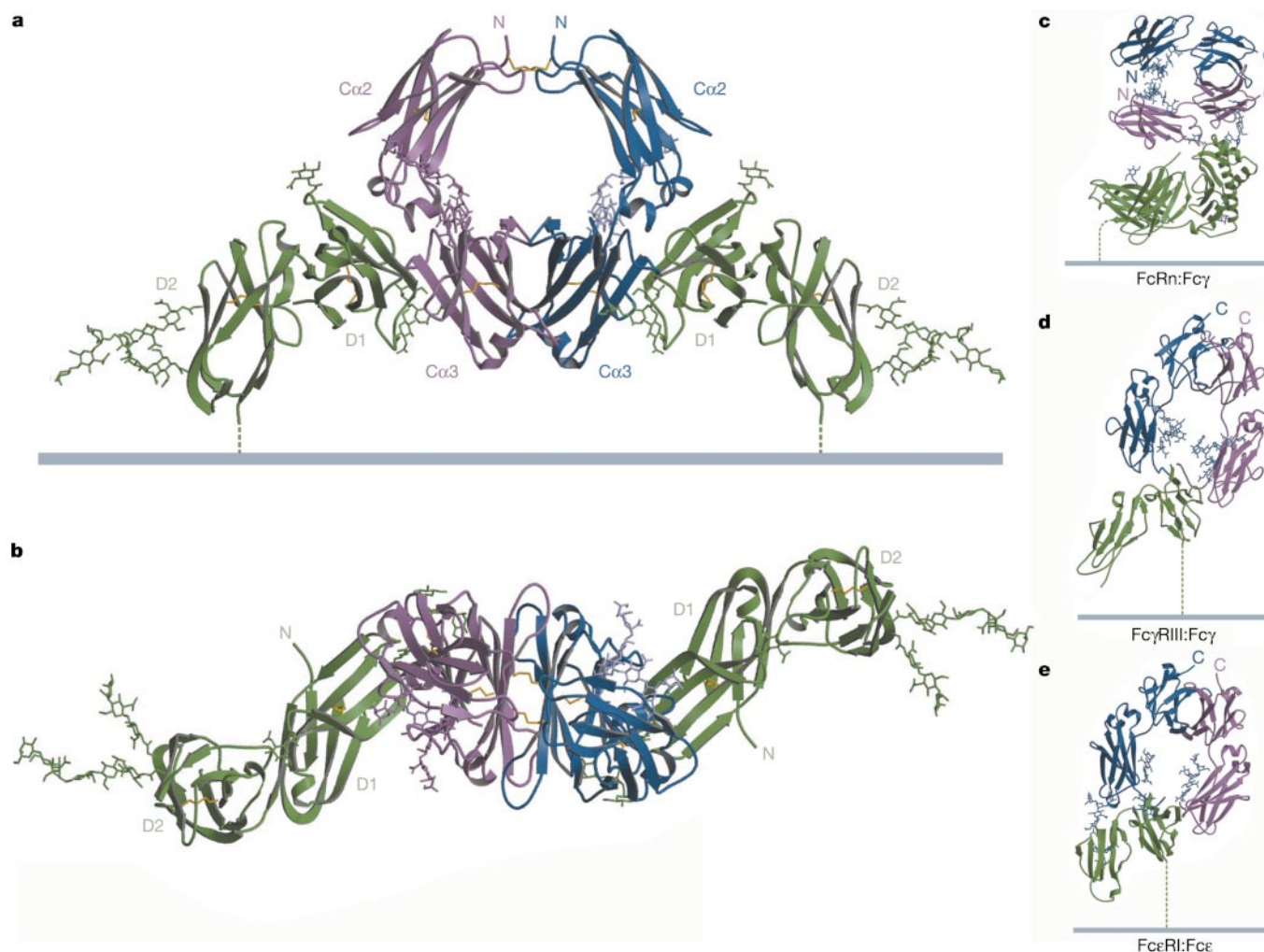


Figure 3 Fc α RI:Fc α structure. **a, b** Ribbon diagrams of the Fc α RI:Fc α complex structure seen from the side (two-fold symmetry axis is vertical) (**a**) or the top (looking down the two-fold symmetry axis) (**b**). The approximate location of the outer surface of the cell membrane is shown as a grey line in the top panel. **c–e** Ribbon diagrams of the

FcRn:Fc γ ¹⁵, Fc γ RIII:Fc γ ¹¹ and Fc ϵ RI:Fc ϵ ¹⁰ complexes showing proposed orientations with reference to the cell membrane. The receptors are green, and the Fcs are blue and magenta.

surface; Supplementary Table 3), and C311 is located only 10 Å away; therefore, bound SC occludes some or all of the FcαRI-binding site(s) in SIgA, apparently preventing binding and activation of FcαRI in the absence of the Mac-1 co-receptor.

FcαRI binds IgA with a 2:1 stoichiometry, as shown here by the co-crystal structure and by independent techniques in solution¹². The 2:1 stoichiometry raises an interesting question: if IgA alone (in the absence of antigen) is able to dimerize FcαRI, what prevents constitutive activation of FcαRI by free IgA? There are several potential explanations. First, the C termini of the two FcαRI

molecules in the FcαRI:Fcα complex are separated by 124 Å, which may be too far apart to register as a receptor dimerization event triggering downstream signalling. Second, the concentration of IgA in serum is typically 20 μM, which is 50- to 100-fold above the K_d for the FcαRI:IgA binding interaction¹². Under such conditions, 1:1 complexes of FcαRI with free IgA would dominate. Due to the moderately fast on- and off-rates of the FcαRI:IgA binding reaction (2×10^5 (M s)⁻¹ and 0.04 s⁻¹, respectively)¹², rapid sampling of free IgA would occur. By contrast, the multivalent nature of interactions between FcαRI and an IgA-immune complex

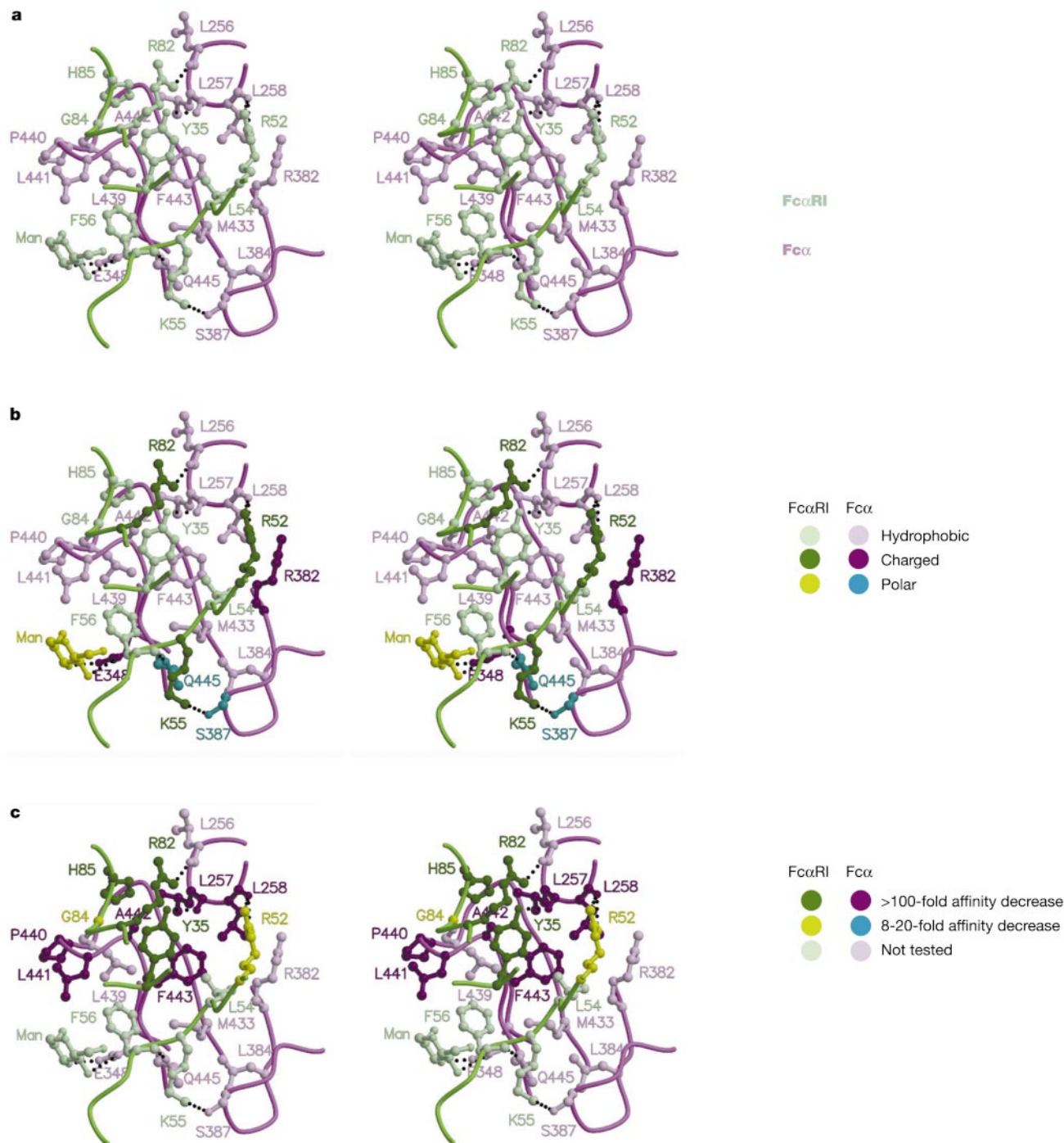


Figure 4 FcαRI:Fcα interface. Stereoviews of residues at the FcαRI:Fcα interface (defined as residues with any non-hydrogen atom within 4 Å of the partner domain). Potential hydrogen bonds are shown as black dotted lines. Residues are colour-coded

according to protein (a), the chemical character of their side chains (b), or their effects on binding affinity when substituted^{13,14,26–28} (c).

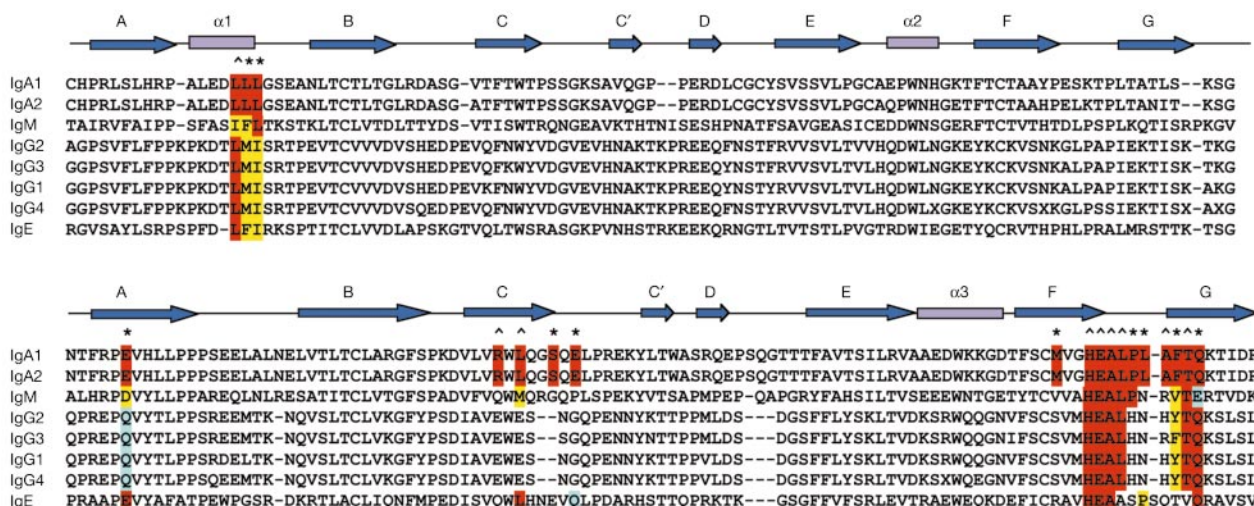


Figure 5 Alignment of antibody Fc sequences. Residues that make up the FcαRI-binding site in IgA1, or exact matches in other isotypes, are in red. Closely matched residues (that is, conserved hydrophobicity or charge) are in yellow. Similar residues (that is, E → Q

substitution) are in blue. The secondary structure of Fcα is illustrated above the sequences. Carets (^) and asterisks (*) denote residues that contribute 0–4% or 5–12%, respectively, of the binding surface.

would lead to a very slow off-rate, allowing stable binding and phagocytosis of an IgA-bound antigen even in the presence of high free IgA concentrations in serum. Finally, the cytoplasmic domain of FcαRI interacts with the cytoskeleton, and this interaction can be modulated by addition of cytokines such as granulocyte–macrophage colony-stimulating factor (GM-CSF)³⁹. GM-CSF stimulates a rapid increase in phagocytosis and affinity of FcαRI for IgA⁴⁰, indicating that cytokine priming may initiate a change in local FcαRI surface density and, therefore, avidity for IgA. These findings suggest that the tethering of FcαRI to the cytoskeleton could spatially restrict FcαRI, preventing clustering until priming of the immune cell with an activating cytokine occurs¹². □

Methods

Crystallization and data collection

The ectodomain of FcαRI and the Fc fragment of IgA1 were expressed and purified as described¹². Crystals of FcαRI were grown in hanging drops containing 20% PEG-8000 and 0.1 M MES, pH 6.0 (FcαRI crystal form I; space group *I*₄), two molecules in the asymmetric unit) or 20% PEG-8000 and 0.1 M sodium phosphate, pH 6.0 (FcαRI crystal form II; space group *I*₂1₂1₂, two molecules per asymmetric unit). Crystals of the FcαRI:Fcα complex were grown from 2:1 molar ratio mixtures of FcαRI and Fcα in 1.7–1.85 M magnesium sulphate heptahydrate and 0.1 M MES, pH 6.5–6.7, in space group *P*₃ with one 2:1 FcαRI:Fcα complex per asymmetric unit. Data processing and scaling was carried out using the HKL package⁴¹ (Supplementary Table 1). Although the complex dataset was only 72% complete in the highest-resolution shell (3.2–3.1 Å), the data were 90% and 99% complete in the next-highest-resolution shells (3.3–3.2 Å and 3.5–3.3 Å, respectively).

Structure determination and refinement

Phases for crystal form I of FcαRI were determined by a combination of multiwavelength anomalous diffraction (MAD) and multiple isomorphous replacement with anomalous scattering (MIRAS) using trimethyl lead acetate and uranyl acetate derivatives. Five lead positions were determined by the program SOLVE⁴² using all three wavelengths in the MAD data set. Five uranium positions were then determined by anomalous difference Fourier analysis. MIRAS phases were calculated using SHARP⁴³, giving an overall figure of merit of 0.24 from 25–3 Å. Solvent flipping with SOLOMON⁴⁴ gave a readily interpretable map containing essentially the entire first FcαRI molecule and most of the second FcαRI molecule. Refinement was carried out using all reflections with *F* > 0 (no sigma cuts were applied) using cycles of torsional, positional and grouped B-refinement with a bulk solvent correction in CNS⁴⁵ with the maximum likelihood Hendrickson-Lattman (MLHL) target using MIRAS phases without solvent flipping. The refinement initially used noncrystallographic symmetry (NCS) constraints; later stages incorporated NCS restraints (300 kcal mol^{−1} Å²), with separate NCS operators for the first and second immunoglobulin-like domains. Regions that differed between the two molecules were not restrained. The final model (*R*_{cryst} = 23.5%, *R*_{free} = 29.1%) contains residues 6–198 of the first molecule; residues 2–111 and 123–193 of the second molecule; six N-acetyl glucosamine residues; and one Tris molecule. The overall B-factors by domain, including all protein atoms in the first FcαRI molecule (and the second molecule), are: D1, 35.0 Å² (33.6 Å²) and D2, 31.7 Å² (61.4 Å²).

The structure of crystal form II of FcαRI was determined by molecular replacement using AMoRe⁴⁶ with coordinates from the form I structure; the molecular replacement search and subsequent rigid-body refinement gave significantly higher correlation coefficients and lower *R*-values in *I*₂1₂1₂ than in *I*₂22. Both molecules in crystal form II were subjected to rigid-body refinement followed by a single round of torsional refinement with a bulk solvent correction and NCS constraints (*R*_{cryst} = 35.2%, *R*_{free} = 39.0%). The starting model had B-factors arbitrarily fixed at 40 Å²; other than the bulk solvent correction, no B-factor refinement was carried out. The resulting model was not refined further owing to the limited resolution of the data.

The FcαRI:Fcα complex was solved by molecular replacement using AMoRe⁴⁶. A polyaniline version of Fcγ (1DN2.pdb³³) and the refined structure of FcαRI (crystal form I) were used as search models, yielding a clear solution in space group *P*₃. The model for a half-complex (one Fcα heavy chain and one FcαRI molecule) was subjected to rigid-body refinement followed by torsional, positional and grouped B-factor refinement using all reflections with *F* > 0. We incorporated a bulk solvent correction and used NCS constraints initially and NCS restraints (300 or 200 kcal mol^{−1} Å²) in later rounds, yielding a final model (*R*_{cryst} = 25.1%, *R*_{free} = 28.4%) that includes residues 242–450 of each Fcα heavy chain; residues 6–195 of FcαRI; 20 N-acetyl glucosamine, 20 mannose, two fucose and two sialic acid residues. The overall B-factors by domain for all protein atoms in the half-complex (and the NCS-related half-complex) are: Fcα Co2, 88.5 Å² (89.4 Å²); Fcα Co3, 71.4 Å² (72.3 Å²); FcαRI D1, 77.0 Å² (79.3 Å²); FcαRI D2, 84.0 Å² (82.2 Å²).

For analyses of interdomain angles, contacts and buried surface areas, FcαRI D1 was defined as residues 6–100 and D2 was defined as residues 101–198; Fcα Co2 was defined as residues 242–342 and Co3 was defined as residues 343–450. Interdomain contact residues were defined as residues within 4 Å of another domain and identified using CNS⁴⁵. Potential hydrogen bonds were identified according to established criteria⁴⁷, using the program HBPLUS⁴⁷. Buried surface areas were calculated using AREA from the CCP4 program suite⁴⁴ with a 1.4 Å probe. Angles between domains were calculated using DOM_ANGLE⁴⁸, which determines the angle between the long axes of adjacent domains that are approximated by ellipsoids calculated from the coordinates. Figures were generated using Molscript⁴⁹ and Raster3D⁵⁰.

Received 26 November 2002; accepted 7 April 2003; doi:10.1038/nature01685.

Published online 21 May 2003.

- Monteiro, R. C. & Van De Winkel, J. G. IgA Fc receptors. *Annu. Rev. Immunol.* **21**, 177–204 (2003).
- Brandtzaeg, P. *et al.* Regional specialization in the mucosal immune system: what happens in the microcompartments? *Immunol. Today* **20**, 141–151 (1999).
- Norderhaug, I. N., Johansen, F. E., Schjerve, H. & Brandtzaeg, P. Regulation of the formation and external transport of secretory immunoglobulins. *Crit. Rev. Immunol.* **19**, 481–508 (1999).
- Monteiro, R. C., Kubagawa, H. & Cooper, M. D. Cellular distribution, regulation, and biochemical nature of an Fcα receptor in humans. *J. Exp. Med.* **171**, 597–613 (1990).
- Vidarsson, G. *et al.* Activity of human IgG and IgA subclasses in immune defense against *Neisseria meningitidis* serogroup B. *J. Immunol.* **166**, 6250–6256 (2001).
- van Egmond, M. *et al.* FcαRI-positive liver Kupffer cells: reappraisal of the function of immunoglobulin A in immunity. *Nature Med.* **6**, 680–685 (2000).
- van Spriel, A. B. *et al.* Mac-1 (CD11b/CD18) is essential for Fc receptor-mediated neutrophil cytotoxicity and immunologic synapse formation. *Blood* **97**, 2478–2486 (2001).
- Motegi, Y. & Kita, H. Interaction with secretory component stimulates effector functions of human eosinophils but not of neutrophils. *J. Immunol.* **161**, 4340–4346 (1998).
- Wende, H., Colonna, M., Ziegler, A. & Volz, A. Organization of the leukocyte receptor cluster (LRC) on human chromosome 19q13.4. *Mamm. Genome* **10**, 154–160 (1999).
- Garman, S. C., Wurzburg, B. A., Tarchevskaya, S. S., Kinet, J. P. & Jardetzky, T. S. Structure of the Fc fragment of human IgE bound to its high-affinity receptor FcεRIα. *Nature* **406**, 259–266 (2000).

11. Sondermann, P., Huber, R., Oosthuizen, V. & Jacob, U. The 3.2-Å crystal structure of the human IgG1 Fc fragment-FcγRIII complex. *Nature* **406**, 267–273 (2000).
12. Herr, A. B., White, C. L., Milburn, C., Wu, C. & Bjorkman, P. J. Bivalent binding of IgA1 to FcαRI suggests a mechanism for cytokine activation of IgA phagocytosis. *J. Mol. Biol.* **327**, 645–657 (2003).
13. Pleass, R. J., Dunlop, J. I., Anderson, C. M. & Woof, J. M. Identification of residues in the CH2/CH3 domain interface of IgA essential for interaction with the human Fcα receptor (FcαR) CD89. *J. Biol. Chem.* **274**, 23508–23514 (1999).
14. Carayannopoulos, L., Hexham, J. M. & Capra, J. D. Localization of the binding site for the monocyte immunoglobulin (Ig) A-Fc receptor (CD89) to the domain boundary between Cα2 and Cα3 in human IgA1. *J. Exp. Med.* **183**, 1579–1586 (1996).
15. Martin, W. L., West, A. P. Jr, Gan, L. & Bjorkman, P. J. Crystal structure at 2.8 Å of an FcRn/heterodimeric Fc complex: mechanism of pH-dependent binding. *Mol. Cell* **7**, 867–877 (2001).
16. Chapman, T. L., Heikema, A. P., West, A. P. & Bjorkman, P. J. Crystal structure and ligand binding properties of the D1D2 region of the inhibitory receptor LIR-1 (ILT2). *Immunity* **13**, 727–736 (2000).
17. Garman, S. C., Kinet, J. P. & Jardetzky, T. S. Crystal structure of the human high-affinity IgE receptor. *Cell* **95**, 951–961 (1998).
18. Maxwell, K. F. *et al.* Crystal structure of the human leukocyte Fc receptor, FcγRIIa. *Nature Struct. Biol.* **6**, 437–442 (1999).
19. Sondermann, P., Huber, R. & Jacob, U. Crystal structure of the soluble form of the human Fcγ-receptor IIb: a new member of the immunoglobulin superfamily at 1.7 Å resolution. *EMBO J.* **18**, 1095–1103 (1999).
20. Fan, Q. R. *et al.* Structure of the inhibitory receptor for human natural killer cells resembles haematopoietic receptors. *Nature* **389**, 96–100 (1997).
21. Huber, R., Deisenhofer, J., Colman, P. M., Matsushima, M. & Palm, W. Crystallographic structure studies of an IgG molecule and an Fc fragment. *Nature* **264**, 415–420 (1976).
22. Wan, T. *et al.* The crystal structure of IgE Fc reveals an asymmetrically bent conformation. *Nature Immunol.* **3**, 681–686 (2002).
23. Wurzburg, B. A., Garman, S. C. & Jardetzky, T. S. Structure of the human IgE-Fc Cε3-Cε4 reveals conformational flexibility in the antibody effector domains. *Immunity* **13**, 375–385 (2000).
24. Putnam, F. W., Liu, Y. S. & Low, T. L. Primary structure of a human IgA1 immunoglobulin. IV. Streptococcal IgA1 protease, digestion, Fab and Fc fragments, and the complete amino acid sequence of the α1 heavy chain. *J. Biol. Chem.* **254**, 2865–2874 (1979).
25. Yang, C., Kratzin, H., Gotz, H. & Hilschmann, N. Rule of antibody structure. Primary structure of a human monoclonal IgA-immunoglobulin (myeloma protein Tro). VII. Purification and characterization of the disulfide bridges. *Hoppe-Seyler's Z. Physiol. Chem.* **360**, 1919–1940 (1979).
26. Morton, H. C. *et al.* Immunoglobulin-binding sites of human FcαRI (CD89) and bovine Fcγ2R are located in their membrane-distal extracellular domains. *J. Exp. Med.* **189**, 1715–1722 (1999).
27. Wines, B. D. *et al.* Identification of residues in the first domain of human Fcα receptor essential for interaction with IgA. *J. Immunol.* **162**, 2146–2153 (1999).
28. Wines, B. D., Sardjono, C. T., Trist, H. M., Lay, C. S. & Hogarth, P. M. The interaction of FcαRI with IgA and its implications for ligand binding by immunoreceptors of the leukocyte receptor cluster. *J. Immunol.* **166**, 1781–1789 (2001).
29. Garman, S. C., Sechi, S., Kinet, J. P. & Jardetzky, T. S. The analysis of the human high affinity IgE receptor FcεRIα from multiple crystal forms. *J. Mol. Biol.* **311**, 1049–1062 (2001).
30. Deisenhofer, J. Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from *Staphylococcus aureus* at 2.9- and 2.8-Å resolution. *Biochemistry* **20**, 2361–2370 (1981).
31. Sauer-Eriksson, A. E., Kleywegt, G. J., Uhlen, M. & Jones, T. A. Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG. *Structure* **3**, 265–278 (1995).
32. Corper, A. L. *et al.* Structure of human IgM rheumatoid factor Fab bound to its autoantigen IgG Fc reveals a novel topology of antibody–antigen interaction. *Nature Struct. Biol.* **4**, 374–381 (1997).
33. DeLano, W. L., Ultsch, M. H., de Vos, A. M. & Wells, J. A. Convergent solutions to binding at a protein–protein interface. *Science* **287**, 1279–1283 (2000).
34. Saphire, E. O. *et al.* Contrasting IgG structures reveal extreme asymmetry and flexibility. *J. Mol. Biol.* **319**, 9–18 (2002).
35. Shogren, R., Gerken, T. A. & Jentoft, N. Role of glycosylation on the conformation and chain dimensions of O-linked glycoproteins: light-scattering studies of ovine submaxillary mucin. *Biochemistry* **28**, 5525–5536 (1989).
36. Iikura, M. *et al.* Secretory IgA induces degranulation of IL-3-primed basophils. *J. Immunol.* **161**, 1510–1515 (1998).
37. Hexham, J. M. *et al.* A human immunoglobulin (Ig)A Cα3 domain motif directs polymeric Ig receptor-mediated secretion. *J. Exp. Med.* **189**, 747–752 (1999).
38. White, K. D. & Capra, J. D. Targeting mucosal sites by polymeric immunoglobulin receptor-directed peptides. *J. Exp. Med.* **196**, 551–555 (2002).
39. Bracke, M., Lammers, J. W., Coffey, P. J. & Koenderman, L. Cytokine-induced inside-out activation of FcαR (CD89) is mediated by a single serine residue (S263) in the intracellular domain of the receptor. *Blood* **97**, 3478–3483 (2001).
40. Weisbart, R. H., Kacena, A., Schuh, A. & Golde, D. W. GM-CSF induces human neutrophil IgA-mediated phagocytosis by an IgA Fc receptor activation mechanism. *Nature* **332**, 647–648 (1988).
41. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
42. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
43. De La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
44. Collaborative Computational Project No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
45. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
46. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
47. McDonald, I. K. & Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793 (1994).
48. Su, X. D. *et al.* Crystal structure of hemolin: a horseshoe shape with implications for homophilic adhesion. *Science* **281**, 991–995 (1998).
49. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
50. Merritt, E. A. & Murphy, M. E. P. Raster3D Version 2.0, a program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873 (1994).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank E. R. Sprague for help with data collection; E. R. Sprague, L. M. Thomas, A. P. West and K. Locher for discussions; P. M. Snow and the Caltech Protein Expression Facility for FcαRI expression; C. L. White for initial FcαRI crystallization trials; staff of the Stanford Synchrotron Radiation Laboratory for technical support; and G. Waksman, M. J. Bennett, W. L. Martin and members of the Bjorkman laboratory for comments on the manuscript. This work was supported by funds from the Damon Runyon Cancer Research Foundation (to A.B.H.), the Howard Hughes Medical Institute (to P.J.B.) and the Ralph M. Parsons Foundation for computational support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.J.B. (bjorkman@caltech.edu). Coordinates have been deposited with the Protein Data Bank under accession codes 1OVZ and 1OWO.

An extragalactic supernebula confined by gravity

J. L. Turner*, S. C. Beck†, L. P. Crosthwaite*‡, J. E. Larkin*, I. S. McLean* & D. S. Meier*§

* Department of Physics and Astronomy, UCLA, Los Angeles, California 90095-1562, USA

† School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Ramat Aviv, Israel

‡ Astute Networks, 16868 Via Del Campo Court, Suite 200, San Diego, California 92127, USA

§ Department of Astronomy, 1002 West Green Street, University of Illinois, Urbana, Illinois 61801, USA

Little is known about the origins of globular clusters, which contain hundreds of thousands of stars in a volume only a few light years across. Radiation pressure and winds from luminous young stars should disperse the star-forming gas and disrupt the formation of the cluster. Globular clusters in our Galaxy cannot provide answers; they are billions of years old. Here we report the measurement of infrared hydrogen recombination lines from a young, forming super star cluster in the dwarf galaxy NGC5253. The lines arise in gas heated by a cluster of about one million stars, including 4,000–6,000 massive, hot ‘O’ stars^{1,2}. It is so young that it is still enshrouded in gas and dust, hidden from optical view^{1,3–5}. The gases within the cluster seem bound by gravity, which may explain why the windy and luminous O stars have not yet blown away those gases. Young clusters in ‘starbursting’ galaxies in the local and distant Universe may also be gravitationally confined and cloaked from view.

NGC5253 is host to hundreds of large star clusters⁶, including dozens of extremely bright, young super star clusters⁷ (SSCs). Only a few million years old^{8,9}, these clusters are found in the central ‘starburst’⁹ region of the nearby (3.8 Mpc) galaxy. Subarcsecond radio and infrared imaging reveal^{3–5} a bright ‘supernebula’ within the starburst, optically invisible, probably powered by a young SSC^{1,2}. The luminosity required to ionize the supernebula is $(0.8–1.2) \times 10^9 L_{\odot}$, where the subscript $\odot$ refers to the Sun, within a 1–2 pc (ref. 1) region—perhaps the most concentrated star-forming luminosity known.

To verify the nature of the supernebula and to study its dynamics, we observed the Brackett α and γ recombination lines of hydrogen at 4.05 μm and 2.17 μm using the NIRSPEC¹⁰ on the Keck Telescope on 11 March 2000. Spectra were taken through a $0.579'' \times 24''$ slit at a spectral resolution of $R \approx 25,000$, or about 12 km s^{-1} . SCAM, a 256×256 array camera within NIRSPEC, simultaneously imaged the slit location on the galaxy at the K band (2.2 μm), allowing us to pinpoint where the spectra were taken. The seeing was $0.55''–0.8''$.

The 2.2 μm broadband image of NGC5253 reveals hundreds of bright infrared star clusters, shown in Figs 1 and 2. From their optical colours, these SSCs are estimated to be only $\sim 2–50$ Myr in age^{8,9}. We find that the brightest infrared source does not coincide with any of the optical clusters. The brightest 2.2- μm source is offset by $\sim 0.3'' \pm 0.1''$ to the northwest of the youngest⁸ optical source. This visually obscured K-band source is the location of the strong Brackett line emission that we observe from the supernebula.

Figure 2 shows the image of the slit position with the strongest Brackett line emission and the corresponding Brackett α echellogram. Line plots of both Brackett lines are shown in Fig. 3. Continuum-subtracted line fluxes are $S_{\text{Br}\alpha}^{\text{obs}} = (3.4 \pm 1) \times 10^{-16} \text{ W m}^{-2}$ and $S_{\text{Br}\gamma}^{\text{obs}} = (3.9 \pm 1) \times 10^{-17} \text{ W m}^{-2}$ for a region within $\sim 3''$ of the continuum peak. For comparison, ref. 11 measures $S_{\text{Br}\alpha}^{\text{obs}} = (7.0 \pm 1) \times 10^{-16} \text{ W m}^{-2}$ and $S_{\text{Br}\gamma}^{\text{obs}} = (1.5 \pm 1) \times 10^{-16} \text{ W m}^{-2}$ in a $10'' \times 20''$ aperture; as much as half of their Br

α flux may be contributed by an unsubtracted continuum. The large aperture flux at the K band appears to be dominated by stellar emission, while much of the L' flux is from nebular dust. Our photometry and mapping confirm that most, if not all, of the Br α and at least 30% of the Br γ emission in NGC5253 emerges from a source coincident with the brightest K-band continuum source.

Having both Brackett lines allows us to correct the line fluxes for extinction and obtain a good estimate of the true, extinction-free recombination line flux, and thus the total ionizing flux. For an intrinsic $S_{\text{Br}\alpha}^{\text{obs}}/S_{\text{Br}\gamma}^{\text{obs}}$ flux ratio of 2.8 (ref. 12), a temperature of 12,000 K (ref. 13), and a Rieke and Lebofsky¹⁴ extinction law, we obtain extinctions of $A_{\alpha} = 0.8 \text{ mag}$ at 4.05 μm and $A_{\gamma} = 2.0 \text{ mag}$ at 2.17 μm . The corresponding optical extinction is $A_v = 18 \text{ mag}$. Optical Balmer recombination lines give $A_v = 3$ (ref. 13). The apparent contradiction between the low optical extinction and higher infrared extinction is resolved if the extinction is high and internal to the nebula. The extinction-corrected Brackett line fluxes are $S_{\text{Br}\alpha}^{\text{obs}} = 7.1 \times 10^{-16} \text{ W m}^{-2}$ and $S_{\text{Br}\gamma}^{\text{obs}} = 2.5 \times 10^{-16} \text{ W m}^{-2}$ for the inner $\sim 1.5''$. These fluxes predict a 15-GHz free-free flux of $24 \pm 8 \text{ mJy}$: the observed 15 GHz flux in this region, corrected for opacity², is $24 \pm 3 \text{ mJy}$ (ref. 4). The Lyman continuum rate required to maintain the ionization of the supernebula is $N_{\text{Ly}\alpha} = (4 \pm 1) \times 10^{52} \text{ s}^{-1}$, equivalent to 4,000 O7 stars. This agrees with centimetre-wave and millimetre-wave free-free fluxes^{1,2,4} and radio recombination lines¹⁵ (Owens Valley millimetre-wave fluxes² give a value of 6,000 O7 stars from a larger aperture). Both the excellent agreement of the Brackett line fluxes and radio fluxes and the observed compactness of the emission argue that the Brackett line emission arises from the radio ‘supernebula’.

The velocity information presented here is new; these spectra

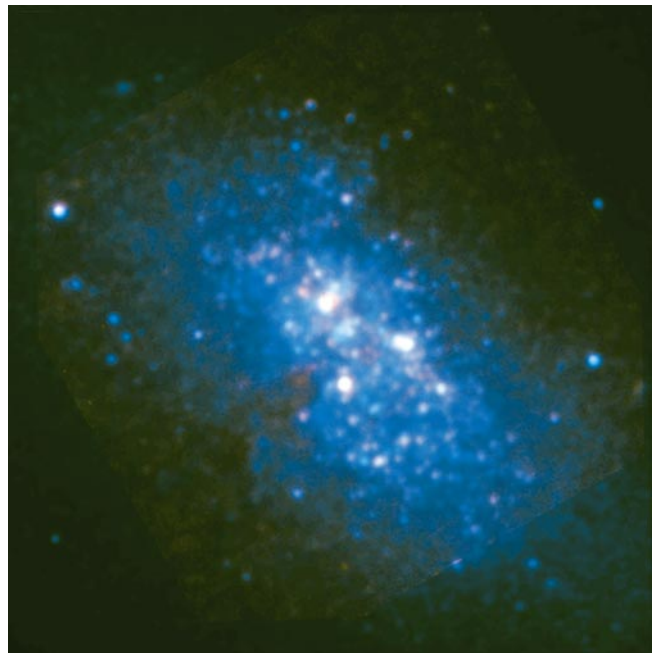


Figure 1 Infrared–optical colour view of the young SSCs of NGC5253. The $\lambda = 2.2 \mu\text{m}$ image (red channel) was constructed from SCAM images made during the night. The infrared image is combined with an optical image from the Hubble Space Telescope (HST) (blue and green channels). Seeing was $\sim 0.7–0.8''$ for the infrared image; the HST image was convolved to match. The brightest K-band source appears as an extended red source to the north of the dust lane. A gaussian fit to this source gives a size (FWHM) of $1.10'' \times 0.9''$, position angle 36° . From smaller clusters we estimate the point spread function for the SCAM image to be $0.75'' \pm 0.2''$. If the K-band source is gaussian, this would imply a source size of $\sim 0.8'' \times 0.5''$, along position angle 40° . This size is uncertain because of variable seeing. The entire SCAM image is $46''$ square. The orientation of the image is east to the left, north up.

probe the dynamics of the nebula at high spatial and spectral resolution. The Brackett line profiles were fitted with gaussians, with line centroids at $v_{\text{LSR}} = 377 \pm 2 \text{ km s}^{-1}$ and full-widths at half-maxima (FWHMs) of $76 \pm 1 \text{ km s}^{-1}$, consistent with $\text{H}\alpha$ (ref. 16) and radio recombination lines¹⁵. For gas at 12,000 K, this linewidth is supersonic. Supersonic gas motions are expected in nebulae: in addition to nebular expansion there is also the interaction of the nebular gas with winds from the massive stars in the cluster. For comparison, compact H II regions in our Galaxy around small groups of O stars have linewidths of up to 64 km s^{-1} (ref. 17); nebulae around individual O stars in the Wolf–Rayet phase can have linewidths of $50\text{--}200 \text{ km s}^{-1}$ (ref. 18).

The first implication of the linewidth of 75 km s^{-1} for the ‘supernebula’ is that if the nebula were actually expanding at the implied speed of $\sim 38 \text{ km s}^{-1}$, then the mean radius of $\sim 0.7 \text{ pc}$ (ref. 1) of the nebula implies an implausibly short dynamical age of 7,000 years. Dynamical ages of compact H II regions in our Galaxy are generally too short to explain the numbers of observed nebulae²¹. It is thought that confinement mechanisms such as the pressure of dense molecular clouds are responsible for the retardation of their expansion. This could lengthen the lifetime of the supernebula phase, if there were molecular gas nearby, which is as yet undetected².

The more unusual implication of the linewidth is that although they are supersonic, these nebular lines are remarkably narrow given the size of the nebula and the high luminosity of its embedded star cluster. Brackett line, radio and infrared fluxes all require $L_{\text{OB}} \approx (0.8\text{--}1.2) \times 10^9 L_{\odot}$ for excitation of the nebula. For a cluster with a Salpeter initial mass function (IMF) and a lower mass cut-off of $1 M_{\odot}$, the mass in stars corresponding to this luminosity is $(5\text{--}7) \times 10^5 M_{\odot}$; if the IMF extends down to stars less than $1 M_{\odot}$, the cluster mass may reach $10^6 M_{\odot}$. The size of the radio nebula is well-determined, and should be the same as the

Brackett size because both scale with emission measure; VLA images show that the nebula is $0.9 \text{ pc} \times 1.8 \text{ pc}$ in size ($0.05'' \times 0.1''$, $\pm 0.02''$) (ref. 1). We assume that the star cluster exciting the nebula lies within the supernebula, otherwise the implied excitation luminosity would be higher than the total observed IRAS luminosity of the entire galaxy, $1.8 \times 10^9 L_{\odot}$ (ref. 5). For a radius of $0.5\text{--}0.9 \text{ pc}$, the escape velocity is $\sim 25\text{--}30 \text{ km s}^{-1}$ for a cluster of O stars alone; $v_{\text{esc}} \approx 50\text{--}70 \text{ km s}^{-1}$ for an IMF cut-off at $1 M_{\odot}$; and $v_{\text{esc}} \approx 85\text{--}110 \text{ km s}^{-1}$ for a cluster IMF extending below $1 M_{\odot}$. Gravity must play a significant role in the dynamics of this nebula, slowing its expansion, if not halting it. Virial linewidths are only slightly smaller than v_{esc} , so the nebula could be in gravitational equilibrium.

If the ‘supernebula’ stage of SSC formation should prove to be common and long-lived, it may have implications for the detection of Lyman α and other ultraviolet lines in star-forming galaxies. There is evidence that Ly α is detectable primarily in galaxies in which the potential absorbing gas is Doppler-shifted out of the path of the Ly α photons²⁰. This would be most likely in galaxies with superwinds and superbubbles, phenomena which can be caused by large concentrations of hot and windy O stars²¹. If a significant fraction of the lifetime of the ionizing phase of an SSC is spent in a confined, dust-enshrouded state such as the supernebula, it is likely that these winds will develop only late in the formation process of the cluster, perhaps at the occurrence of the first supernovae. By that time the ionizing flux of the cluster is already declining. SSCs may be an important mode of star formation in the early Universe. Ly α searches for primeval galaxies may therefore fall far short of detecting the true star formation rate, as suggested by observations^{20,22–24}.

To put the supernebula in context, we can compare it to a more traditional nebula, 30 Doradus in the Large Magellanic Cloud. 30 Dor is a large, luminous nebula ionized by an optically visible star cluster with an estimated age of 10 Myr (ref. 25), $150\text{--}200 \text{ pc}$ in

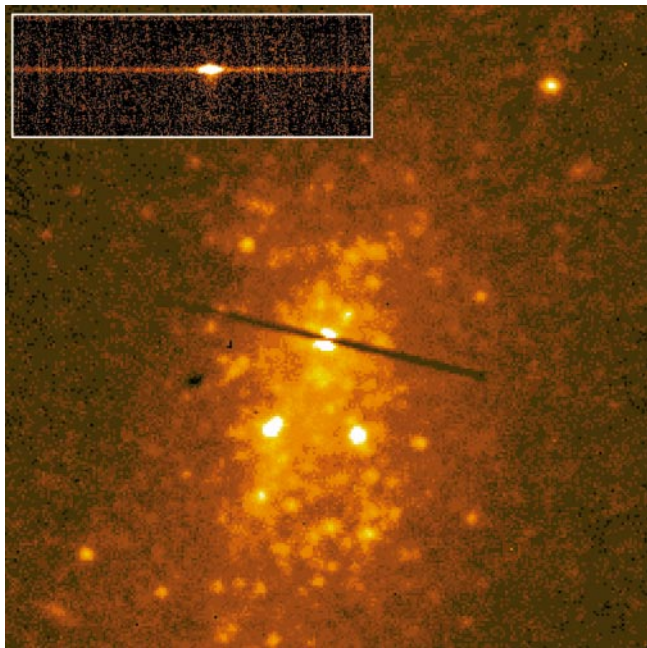


Figure 2 Echellogram and slit position for Brackett spectra of the supernebula. The $46'' \times 46''$ SCAM image shows the position of the $0.579'' \times 24''$ slit on the brightest K-band source in NGC5253. In the inset is the Brackett α echellogram, with frequency/velocity running horizontally and the spatial dimension vertically. The full slit length is $24''$, but only $6''$ is shown in the inset. The orientation of the image is such that north is at an angle of -43° (clockwise) from vertical. The Brackett line emission is less than $1.3''$ in spatial extent on the echellogram, the same as the standard star. Spectra taken at eight other positions showed weak or no emission.

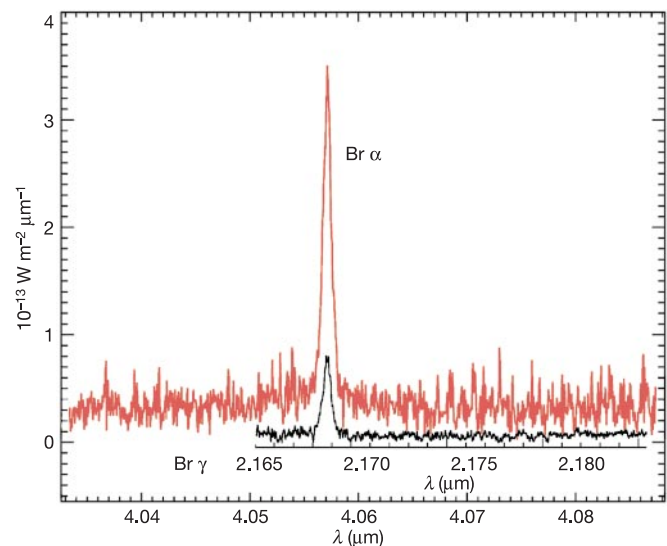


Figure 3 Spectra of Br α and Br γ in NGC5253. The spectra were integrated over the central $3''$ centred on the brightest infrared continuum source. Each grating setting was followed by a calibration A star. On/off slit SCAM images of the standard indicate that 50% of the light entered the slit. Because the Brackett line emission has a spatial extent comparable to the standard star, $\sim 6\text{--}9$ pixels ($0.8''\text{--}1.3''$), calibration using the standard should automatically correct for this effect. We estimate an uncertainty of 30% in the line fluxes due to variability in seeing. Near-infrared continuum emission from the brightest K-band source is so strong that it can be seen even in these highly dispersed spectra. We measure continuum fluxes of $186 \pm 60 \text{ mJy}$ and $12 \pm 4 \text{ mJy}$ at L' and K, respectively, as compared to the 144 mJy and 20 mJy fluxes in these bands observed in ref. 29 with $8\text{--}9''$ apertures.

extent and with a gas density of $20\text{--}100\text{ cm}^{-3}$ (ref. 26). 30 Dor is 200 times larger and $100\text{--}1,000$ times less dense than the supernebula in NGC5253, and its exciting star cluster R136 is $10\text{--}100$ times less massive than the cluster in NGC5253 (ref. 27). The escape velocity for 30 Dor is less than the thermal sound speed of 10 km s^{-1} so its gas motions are determined by winds and turbulence rather than by gravity²⁸. How the supernebula in NGC5253 will free itself from 'gravitational bondage' and evolve into an extended, diffuse nebula like 30 Dor will probably depend on the evolution of its underlying star cluster. □

Received 27 November 2002; accepted 28 April 2003; doi:10.1038/nature01689.

1. Turner, J. L., Beck, S. C. & Ho, P. T. P. The radio supernebula in NGC5253. *Astrophys. J.* **532**, L109–L112 (2000).
2. Meier, D. S., Turner, J. L. & Beck, S. C. Molecular gas and star formation in NGC5253 revisited. *Astron. J.* **124**, 877–885 (2002).
3. Beck, S. C., Turner, J. L., Ho, P. T. P., Kelly, D. & Lacy, J. H. The central star cluster of the star-forming dwarf galaxy NGC5253. *Astrophys. J.* **457**, 610–615 (1996).
4. Turner, J. L., Ho, P. T. P. & Beck, S. C. The radio properties of NGC5253 and its unusual HII regions. *Astron. J.* **116**, 1212–1220 (1998).
5. Gorjian, V., Turner, J. L. & Beck, S. C. Infrared emission from the radio supernebula in NGC5253: A proto-globular cluster? *Astrophys. J.* **554**, L29–L32 (2001).
6. Caldwell, N. & Phillips, M. M. Star formation in NGC5253. *Astrophys. J.* **338**, 789–803 (1989).
7. Gorjian, V. WFPC2 imaging of the starburst galaxy NGC5253. *Astron. J.* **1120**, 1886–1893 (1996).
8. Calzetti, D. *et al.* Dust and recent star formation in the core of NGC5253. *Astron. J.* **114**, 1834–1849 (1997).
9. Tremonti, C. A., Calzetti, D., Leitherer, C. & Heckman, T. M. Star formation in the field and clusters of NGC5253. *Astrophys. J.* **555**, 322–337 (2001).
10. McLean, I. S. *et al.* Performance and results with the NIRSPEC echelle spectrograph on the Keck II Telescope. *Proc. SPIE* **4008**, 1048–1055 (2000).
11. Kawara, K., Nishida, M. & Phillips, M. M. Brackett alpha and gamma observations of starburst and Seyfert galaxies. *Astrophys. J.* **337**, 230–235 (1989).
12. Storey, P. J. & Hummer, D. G. Recombination line intensities for hydrogenic atoms. IV. Total recombination coefficients and machine-readable tables for $Z = 1$ to 8. *Mon. Not. R. Astron. Soc.* **272**, 41–48 (1995).
13. Walsh, J. R. & Roy, J.-R. Optical spectroscopic and abundance mapping of the amorphous galaxy NGC5253. *Mon. Not. R. Astron. Soc.* **239**, 297–324 (1989).
14. Rieke, G. H. & Lebofsky, M. J. The interstellar extinction law for 1 to 13 microns. *Astrophys. J.* **288**, 618–621 (1985).
15. Mohan, N. R., Anantharamaiah, K. R. & Goss, W. M. Very Large Array observations of the H92 α line from NGC5253 and Henize 2–10: Ionized gas around super star clusters. *Astrophys. J.* **557**, 659–670 (2001).
16. Martin, C. L. & Kennicutt, R. C. Jr Soft X-ray emission from NGC5253 and the ionized interstellar medium. *Astrophys. J.* **447**, 171–183 (1995).
17. De Pree, C. G., Wilner, D. J., Goss, W. M., Welch, W. J. & McGrath, E. Ultracompact HII regions in W49N at 500 AU scales: Shells, winds, and the water maser source. *Astrophys. J.* **540**, 308–315 (2000).
18. Chu, Y. H. Ring nebulae around massive stars throughout the HR diagram. In *IAU Symposium 212* (eds van der Hucht, K. A., Herrero, A. & Esteban, C.) (in the press).
19. Wood, D. O. S. & Churchwell, E. Massive stars embedded in molecular clouds: Their population and distribution in the Galaxy. *Astrophys. J.* **340**, 265–272 (1989).
20. Kunth, D. *et al.* HST study of Lyman-alpha emission in star-forming galaxies: the effect of neutral gas flows. *Astron. Astrophys.* **334**, 11–20 (1998).
21. Tenorio-Tagle, G., Silich, S. A., Kunth, D., Terlevich, E. & Terlevich, R. The evolution of superbubbles and the detection of Iya in star-forming galaxies. *Mon. Not. R. Astron. Soc.* **309**, 332–342 (1999).
22. Meier, D. L. & Terlevich, R. Extragalactic HII regions in the UV. Implications for primeval galaxies. *Astrophys. J.* **246**, L109–L113 (1981).
23. Hartmann, L., Huchra, J. P. & Geller, M. J. How to find galaxies at high redshift. *Astrophys. J.* **287**, 487–491 (1984).
24. Steidel, C. C., Giallisco, M., Pettini, M., Dickinson, M. & Adelberger, K. L. Spectroscopic confirmation of a population of normal star-forming galaxies at redshifts $z > 3$. *Astrophys. J.* **462**, L17–L21 (1996).
25. Walborn, N. R. & Barbá, R. H. in *New Views of the Magellanic Clouds* (eds Chu, Y.-H., Suntzeff, N., Hesser, J. & Bohlender, D.) 213 (Astronomical Society of the Pacific, San Francisco, 1999).
26. Mills, B. Y., Turtle, A. J. & Watkinson, A. A radio model of the 30 Doradus region. *Mon. Not. R. Astron. Soc.* **185**, 263–276 (1978).
27. Kennicutt, R. C. Jr & Chu, Y.-H. Giant HII regions and the formation of populous clusters. *Astron. J.* **95**, 720–730 (1988).
28. Melnick, J., Tenorio-Tagle, G. & Terlevich, R. Supersonic gas motion in extragalactic HII regions. *Mon. Not. R. Astron. Soc.* **302**, 677–683 (1999).
29. Moorwood, A. F. M. & Glass, I. S. Infrared emission and star formation in NGC5253. *Astron. Astrophys.* **115**, 84–89 (1982).

Acknowledgements We thank T. Glassman and E. Greisen for assistance with the data and M. Jura for discussions. This research is supported by the US National Science Foundation, the Israel Academy Center for Multi-Wavelength Astronomy, and the Laboratory of Astronomical Imaging at the University of Illinois.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.T. (turner@astro.ucla.edu).

A strong decrease in Saturn's equatorial jet at cloud level

A. Sánchez-Lavega*, S. Pérez-Hoyos*, J. F. Rojas†, R. Hueso* & R. G. French‡

* Departamento Física Aplicada I, Escuela Superior de Ingenieros, Universidad del País Vasco, Alameda Urquijo s/n, 48013 Bilbao, Spain

† Departamento Física Aplicada I, EUITI, Universidad del País Vasco, Plaza Casilla s/n, 48013 Bilbao, Spain

‡ Department of Astronomy, Wellesley College, Wellesley, Massachusetts 02481, USA

The atmospheres of the giant planets Jupiter and Saturn have a puzzling system of zonal (east–west) winds alternating in latitude, with the broad and intense equatorial jets on Saturn having been observed previously to reach a velocity of about 470 m s^{-1} at cloud level¹. Globally, the location and intensity of Jupiter's jets are stable in time to within about ten per cent^{2,3}, but little is known about the stability of Saturn's jet system. The long-term behaviour of these winds is an important discriminator between models for giant-planet circulations^{4–9}. Here we report that Saturn's winds show a large drop in the velocity of the equatorial jet of about 200 m s^{-1} from 1996 to 2002. By contrast, the other measured jets (primarily in the southern hemisphere) appear stable when compared to the Voyager wind profile of 1980–81.

The first precise measurements of Saturn's zonal wind velocity profile were performed during the Voyager encounters in 1980–81 (refs 10, 11). Previous wind measurements, obtained by tracking individual 'spots' in the atmosphere, are scarce, but are in general agreement with the Voyager data¹². Cloud motions measured on Voyager images revealed a strong and broad equatorial jet (peak velocity $\sim 470\text{ m s}^{-1}$, spanning planetographic latitudes $\pm 40^\circ$), twice as broad and four times more intense than that of Jupiter. The equator is also the place where infrequent but violent eruptions (the 'Great White Spots') occur^{13–15}. As Saturn has a large obliquity of 26.7° , and the shadow of the rings produces strong insolation changes in the equatorial region, induced effects during Saturn's year (29.4 terrestrial years) could be expected at cloud level where the winds are measured¹⁶. The zonal jet measurements rely on the detection and tracking of cloud features, and because of the lower contrast, size and number of such features in Saturn than in Jupiter, this can only be performed at present using the Hubble Space Telescope (HST). The Voyager data were obtained during the northern hemisphere spring (1980–81), and HST high-quality images of Saturn are available for the period 1994–2002, during Saturn's southern hemisphere spring and early summer. Both epochs are well placed to look for long-term changes.

The HST images analysed here were obtained between 1996 and 2002 (3–6 days per year) using the Wide Field Planetary Camera 2 (WFPC2) at its maximum spatial resolution. A series of filters from 255 to 1,042 nm were used, including at times a filter isolating the 890-nm methane absorption band (dates and filters for each campaign are given in Supplementary Information). In total, about 100 images were selected, navigated on the planetary limb and photometrically calibrated. Images obtained at wavelengths of 439 nm, 675 nm, 814 nm and 890 nm showed the clouds to have high contrast (enhanced by an 'unsharp masking' processing), and were used for target identification and tracking. Image pairs or triplets obtained with the same filter, separated by a temporal interval ranging from 0.5 to 8 days, were used for the tracking. Following the ring-plane crossing epoch in 1995, Saturn's orbital obliquity and ring projection limited our visibility of the northern hemisphere, so most features were observed from latitudes $+30^\circ$ to

–90°. A total of 343 individual features (most with sizes ranging from ~500 to 2,000 km) were tracked. A sample of Saturn's cloud features observed at different latitudes and with different filters, and their tracked motions in the equator, are shown in Fig. 1.

In Fig. 2 we present our 1996–2002 measurements, and compare them to the Voyager wind profile¹¹. We also include individual spot velocities obtained during 1994–97 from ground-based telescopes¹⁷ (full wind speed data are available as Supplementary Information). First, the most evident and unexpected result is the truncation of the equatorial eastward jet to maximum velocities of ~275 m s⁻¹, a drop of ~200 m s⁻¹ relative to the Voyager value. The deviation from the Voyager profile took place equatorward of latitudes ±17°, although the jet itself is slightly broader, starting to change at latitudes ±25°. Second, there are no apparent changes in the other southern eastward and westward jets, even in the 'ribbon' wave jet at latitude +47° re-observed during 1994–96 (ref. 18). Last, we confirm the existence of a recently discovered south polar jet at –73° latitude¹⁹, making Saturn's zonal jet system highly symmetric.

The HST profile is plotted using features measured with all filters and over the full time period indicated above. We do not find any dependence of wind velocities with wavelength, nor with time, during the observing period. The peak truncation has persisted since ~1994. The velocity dispersion seen in the equatorial region from +5° to –17° (amounting in some places up to ~75 m s⁻¹) is probably real, and reflects intense local motions related to the disturbances in the equatorial area (also observed one month after

the 1990 storm outbreak¹⁵). A detailed analysis of the equatorial cloud morphology showed that this period was characterized by the development of large-scale features (~20,000 km in zonal extent) with abundant internal structure (Fig. 1). This differs from the Voyager era, when the most representative cloud features were series of bright plume-like structures¹¹. However, between both periods, 1980–81 and 1996–2002, a large storm erupted at +5° in 1990 (refs 13–15), disrupting the equatorial region cloud morphology from latitudes +22° to –5°, and extending its activity during 1991 (ref. 20). It was followed by another intense storm centred at +10° in 1994 (ref. 21), and by subsequent similar features during 1995–97 (ref. 17).

In order to determine the altitude of the features we see, we have carried out a radiative transfer analysis of the wavelength dependence of the reflectivity across selected latitudes from the equator to the south pole (from limb to limb), using the 'doubling-adding technique'^{21,19,22}. At –10°, where the main jet speed velocity changes are seen, the model that best fits the reflectivity has a high stratospheric haze (altitude between 1 and 10 mbar), and below it, a thick layer of particles (haze or cloud) with optical depth $\tau = 10$ located between 40 and 300 mbar. Deeper (1.8 bar), a putative ammonia cloud is placed. But the features that we see (that is, the cloud tops, where $\tau \approx 1$ –2) are located in the intermediate thick haze, nominally at pressures ~100 mbar, close to the tropopause level^{23,24}. According to cloud models of the equatorial region at the Voyager epoch²⁵ and before the development of the 1990 storm²⁶, the equatorial haze layer reached $\tau = 1$ in the level ~300 mbar. Our analysis of the 1990 storm placed its cloud tops at 200 mbar (ref. 22).

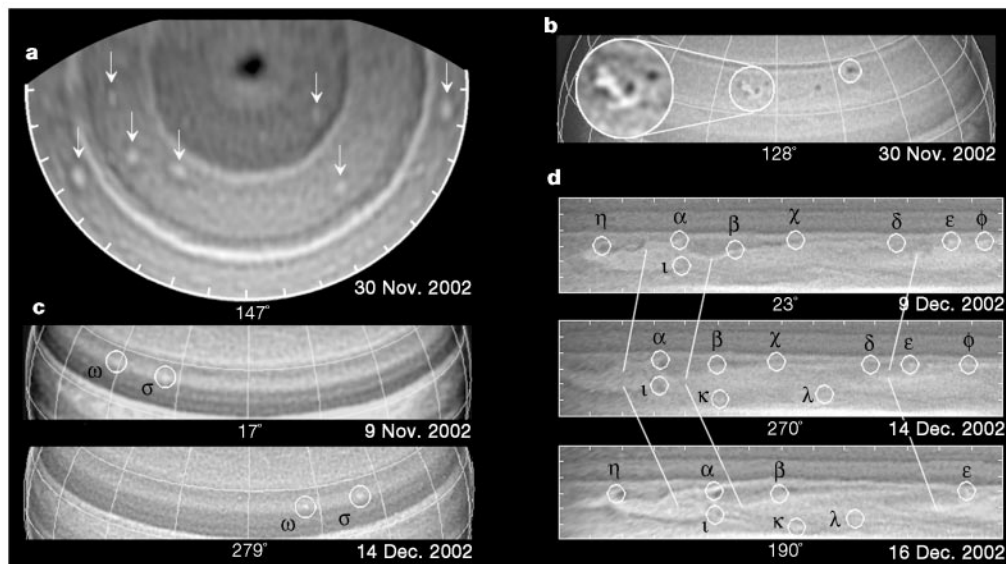


Figure 1 Images of different latitudes of Saturn, showing a representative variety of cloud features tracked to measure the winds. The HST WFPC2 images were obtained at different wavelengths, and in some cases their motions are indicated. The spatial resolution of the camera (PC mode) was 0.0455 arcsec pixel⁻¹ (~260 km pixel⁻¹ at Saturn). **a**, Polar projection. The pole is centred on the dark cap; the arrows mark the bright spots (at a wavelength of 890 nm) at latitudes ranging from –50° to –90° (each tick mark represents 10° in longitudinal extent). **b**, Temperate latitudes at a wavelength of 675 nm (the full latitude circles shown are –35° and –50°). The image shows regularly observed mid-latitude dark spots together with a bright spot (shown magnified in the inset to reveal its structure). **c**, Pair of images showing the zonal translation of two bright features at latitude –29° and at a wavelength of 439 nm. The full latitude circles shown are –17° and –33° (note the change in central meridian longitude between the two

dates). **d**, Cylindrical maps of the equator (from 0° to –30°), showing a variety of features at a wavelength of 890 nm and their zonal translations (note again the changing longitude of the central meridian for each day). Separations between tick marks are 10° in longitude and 5° in latitude. The three white lines identify typical large-scale equatorial features in the three maps. In **b–d**, south is up and east is to the left. Longitudes in the central meridian are based on the System III radio rotation period, and latitudes are planetographic¹¹. The longitudinal translation of each feature in **c** and **d** was converted to drift rate (degrees per day) and then to zonal velocity (see Fig. 2). If the features shown in **d** (first panel) had moved at the Voyager zonal velocities, they would be about 40° to the left of their marked positions in the second panel. This is substantially greater than the measurement uncertainty.

Thus, within the model uncertainties, the equatorial features we see in 1996–2002 are slightly higher than in 1980–81, probably by no more than a scale height (~ 36 km).

The abrupt truncation of the eastward equatorial jet could be the result of a true temporal change in the general circulation due to unknown dynamical processes, or an indirect effect. One possibility is that the cloud systems in 1996–2002, placed at a higher altitude than in 1980–81, moved more slowly, according to the expected decrease in altitude of zonal winds measured in other latitudes²³. Another possibility is that wave or disturbance propagation at cloud level in the equator, relative to the mean zonal flow, is affecting our velocity measurements. However, the historical large-scale equatorial storms (in 1876, 1933 and 1990) moved with speeds ~ 400 m s^{-1} (refs 11, 20), close to the Voyager profile at the central latitude of the storms, indicating that such an effect was not present at least during the initial phase of the storm development.

The picture that emerges from our 1996–2002 measurements, when compared to the Voyager data for 1980–81, is that Saturn's equatorial dynamics and jet structure show high variability. It is the region where most of the large-scale storms form on the planet, and the place where the jet wind speed has changed by 42%, as reported here. Such a large wind speed variability has not been detected in Jupiter, where many more observations on jet stability are avail-

able^{2,3}. Outside the equator, our observations indicate that Saturn's jets appear to be stable in time. Compared to Jupiter, Saturn has fewer jets overall, but they show higher north–south symmetry in latitude and in speed intensity. Saturn receives less sunlight and releases less internal heat than Jupiter, but the sunlight source is more variable owing to the planet's obliquity. For example, the polar areas each alternately fall into darkness during the course of a half Saturn year, and the equator suffers the ring-shadowing insolation variability. All these differences between the planets need to be considered when developing future general circulation models for the jet origin in the giant planets. □

Received 5 February; accepted 11 April 2003; doi:10.1038/nature01653.

- Ingersoll, A. P. Atmospheric dynamics of the outer planets. *Science* **248**, 308–315 (1990).
- García-Melendo, E. & Sánchez-Lavega, A. A study of the stability of Jovian zonal winds from HST images: 1995–2000. *Icarus* **152**, 316–330 (2001).
- Porco, C. C. *et al.* Cassini imaging of Jupiter's atmosphere, satellites and rings. *Science* **299**, 1541–1547 (2003).
- Williams, G. P. Jovian and comparative atmospheric modeling. *Adv. Geophys.* **A 28**, 381–429 (1985).
- Cho, J. Y. K. & Polvani, L. M. The morphogenesis of bands and zonal winds in the atmospheres on the giant outer planets. *Science* **273**, 335–337 (1996).
- Busse, F. H. A simple model of convection in the Jovian atmosphere. *Icarus* **30**, 255–260 (1976).
- Ingersoll, A. P. & Pollard, D. Motion in the interiors and atmospheres of Jupiter and Saturn: Scale analysis, anelastic equations, barotropic stability criterion. *Icarus* **52**, 62–80 (1982).
- Zhang, K. K. & Schubert, G. Teleconvection: Remotely driven thermal convection in rotating stratified spherical layers. *Science* **290**, 1944–1947 (2000).
- Yano, J.-I., Talagrand, O. & Drossart, P. Atmospheric zonal jets in the atmospheres of the giant planets simulated by an idealized two-dimensional deep-fluid model. *Nature* **421**, 36 (2003).
- Ingersoll, A. P., Beebe, R. F., Conrath, B. J. & Hunt, G. E. in *Saturn* (eds Gehrels, T. & Matthews, M. S.) 195–238 (Univ. Arizona Press, Tucson, 1984).
- Sánchez-Lavega, A., Rojas, J. F. & Sada, P. V. Saturn's zonal winds at cloud level. *Icarus* **147**, 405–420 (2000).
- Sánchez-Lavega, A. Motions in Saturn's atmosphere: Observations before Voyager encounters. *Icarus* **49**, 1–16 (1982).
- Sánchez-Lavega, A. *et al.* The Great White Spot and disturbances in Saturn's equatorial atmosphere during 1990. *Nature* **353**, 397–401 (1991).
- Beebe, R. F., Barnet, C., Sada, P. V. & Murrell, A. S. The onset and growth of the 1990 equatorial disturbance on Saturn. *Icarus* **95**, 163–172 (1992).
- Barnet, C. D., Westphal, J. A., Beebe, R. F. & Huber, L. F. Hubble Space Telescope observations of the 1990 equatorial disturbance on Saturn: Zonal winds and central meridian albedos. *Icarus* **100**, 499–511 (1992).
- Barnet, C. D., Beebe, R. F. & Conrath, B. J. A seasonal radiative–dynamic model of Saturn's troposphere. *Icarus* **98**, 94–107 (1992).
- Sánchez-Lavega, A., Lecacheux, J., Colas, F., Rojas, J. F. & Gomez, J. M. Discrete cloud activity in Saturn's equator during 1995, 1996 and 1997. *Planet. Space Sci.* **47**, 1277–1283 (1999).
- Sánchez-Lavega, A. Observations of Saturn's ribbon wave 14 years after its discovery. *Icarus* **158**, 272–275 (2002).
- Sánchez-Lavega, A., Pérez-Hoyos, S., Acarreta, J. R. & French, R. G. No hexagonal wave around Saturn's southern pole. *Icarus* **160**, 216–219 (2002).
- Sánchez-Lavega, A., Lecacheux, J., Colas, F. & Laques, P. Temporal behavior of cloud morphologies and motions in Saturn's atmosphere. *J. Geophys. Res.* **98**, 18857–18872 (1993).
- Sánchez-Lavega, A. *et al.* Large-scale storms in Saturn's atmosphere during 1994. *Science* **271**, 631–634 (1996).
- Acarreta, J. R. & Sánchez-Lavega, A. Vertical cloud structure in Saturn's 1990 equatorial storm. *Icarus* **137**, 24–33 (1999).
- Conrath, B. J. & Pirraglia, J. A. Thermal structure of Saturn from Voyager infrared measurements: Implications for atmospheric dynamics. *Icarus* **53**, 286–292 (1983).
- Lindal, G. F., Sweetnam, D. N. & Eshleman, V. R. The atmosphere of Saturn: An analysis of the Voyager radio occultation measurements. *Astron. J.* **90**, 1136–1146 (1985).
- Tomasko, M. G., West, R. A., Orton, G. S. & Tejfel, V. G. in *Saturn* (eds Gehrels, T. & Matthews, M. S.) 150–194 (Univ. Arizona Press, Tucson, 1994).
- Karkoschka, E. & Tomasko, M. G. Saturn's upper troposphere 1986–1989. *Icarus* **97**, 161–181 (1992).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Flasar for suggestions. The NASA/ESA Hubble Space Telescope is operated by AURA under NASA contract. A.S.-L., S.P.-H. and J.F.R. were supported by the MCYT Plan Nacional de Astronomía y Astrofísica and Grupos UPV. S.P.-H. acknowledges a PhD fellowship from the Spanish MECED, and R.H. a post-doctoral fellowship from Gobierno Vasco. R.G.F. was supported in part by NASA's Planetary Geology and Geophysics Program and an STSCI grant.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.S.-L. (wupsalaa@bi.ehu.es).

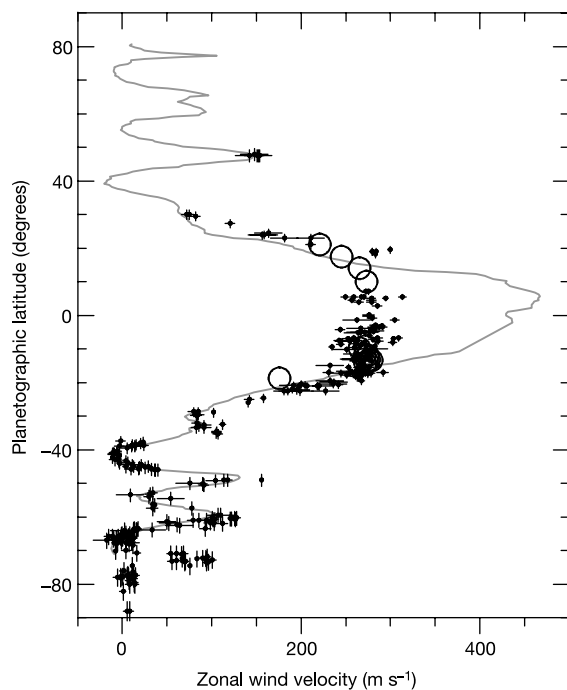


Figure 2 Saturn's zonal (east–west) wind velocity profile at cloud level as a function of latitude. Data are shown for two different epochs (1980–81 and 1995–2002). The continuous line shows the averaged Voyager 1 and 2 profiles, as measured during 1980–81 (ref. 11). The large open circles represent individual measurements obtained from 1995 to 1997 by tracking large features using ground-based telescopes¹⁷. The small black dots are the individual velocity measurements obtained between 1996 and 2002 using HST images as reported in the text, with the error for each individual measurement added as a thin line. The wind speeds are determined relative to the radio rotation period of 10 h 39 min 24 s, assumed to be that of the interior. The main uncertainties in these data arise from image navigation (planetary limb fit), target misidentification, and cursor pointing on the feature. The errors in latitude positioning range from $\pm 0.5^\circ$ in the equator to $\pm 1.5^\circ$ at the poles. Wind speed precision is of the order of target size/tracking time; at our typical target size (500–2,000 km) and tracking times (minimum 0.5 d, maximum 60 d), this gives values ranging from ± 1 m s^{-1} to ± 15 m s^{-1} (see Supplementary Information for the full data for each point used in this figure).

Universal alignment of hydrogen levels in semiconductors, insulators and solutions

Chris G. Van de Walle* & J. Neugebauer†

* Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, California 94304, USA

† Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14195 Berlin-Dahlem, Germany

Hydrogen strongly affects the electronic and structural properties of many materials. It can bind to defects or to other impurities, often eliminating their electrical activity; this effect of defect passivation is crucial to the performance of many photovoltaic and electronic devices^{1,2}. A fuller understanding of hydrogen in solids is required to support development of improved hydrogen-storage systems³, proton-exchange membranes for fuel cells, and high-permittivity dielectrics for integrated circuits. In chemistry and in biological systems, there have also been many efforts to correlate proton affinity and deprotonation with host properties⁴. Here we report a systematic theoretical study (based on *ab initio* methods) of hydrogen in a wide range of hosts, which reveals the existence of a universal alignment for the electronic transition level of hydrogen in semiconductors, insulators and even aqueous solutions. This alignment allows the prediction of the electrical activity of hydrogen in any host material once some basic information about the band structure of that host is known. We present a physical explanation that connects the behaviour of hydrogen to the line-up of electronic band structures at heterojunctions.

Hydrogen exhibits qualitatively different behaviour depending on the host into which it is introduced. As an isolated interstitial impurity, hydrogen can occupy a number of different lattice sites and significantly modify the host structure, to the point of breaking host-atom bonds. It can act as either a donor (H^+) or an acceptor (H^-)—that is, it is amphoteric—and has mostly been found to counteract the conductivity caused by extrinsic dopants^{1,5}. In some cases, however, hydrogen has been found to act as a source of conductivity^{6–8}. This highly host-dependent behaviour raises questions about the underlying physical mechanisms. Some experimental information about interstitial hydrogen is available^{1,5,7}, but to investigate the trends that are the subject of the present study we have relied on *ab initio* methods, which allow us to screen systematically a large range of systems and properties to identify underlying relations and hidden rules. Our calculations for a representative set of semiconductors and insulators allow us to establish a connection between hydrogen's electrical characteristics

and the line-up of band structures of the host materials on an absolute energy scale. This line-up is a physics problem in its own right, with important technological applications⁹. We will demonstrate that the alignment of the hydrogen level extends to aqueous solutions, highlighting its true universality.

Our approach is based on density-functional theory (DFT) within the local density approximation (LDA) and the pseudopotential-plane-wave method^{10,11}. The key quantities that determine the properties of hydrogen are (1) the formation energy, that is, the energy needed to incorporate H in the host, and (2) the electronic transition level, which defines the electrical behaviour. We obtain the formation energy of interstitial H in charge state q (where $q = -1, 0$ or $+1$) by placing H in a volume of host material (a periodically repeated supercell), calculating the total energy $E_{\text{tot}}(H^q)$ of this structure, and subtracting the energy $E_{\text{tot}}(\text{bulk})$ of a corresponding volume of pure host material:

$$E^f(H^q) = E_{\text{tot}}(H^q) - E_{\text{tot}}(\text{bulk}) - 1/2E_{\text{tot}}(H_2) + qE_F \quad (1)$$

Here the reference for the hydrogen energy is given by an H_2 molecule at temperature $T = 0$. The last term in the formation energy accounts for the fact that H^+ donates an electron, and H^- accepts an electron; the energy of the reservoir with which these electrons are exchanged is the electron chemical potential or Fermi level, E_F . The transition level $\epsilon(q+1/q)$ between charge states $q+1$ and q is defined as the Fermi-level position for which the formation energies of these charge states are equal.

Figure 1a illustrates these concepts with the example of H in GaN. H^0 is never the lowest-energy state (characteristic of a 'negative-U' system⁵), causing the donor level $\epsilon(+/-)$ to lie above the acceptor level $\epsilon(0/-)$. When E_F moves through the bandgap, the stable charge state thus changes directly from positive (for E_F below 2.4 eV) to negative (for E_F above 2.4 eV). This implies that in p-type GaN (E_F close to the valence-band maximum, VBM) H^+ is favoured, whereas in n-type GaN (E_F close to the conduction-band minimum, CBM) H^- is stable, providing the basis for hydrogen's tendency to counteract the prevailing conductivity. The Fermi-level position where the positive and negative charge states are equal in energy is labelled $\epsilon(+/-)$, and plays a crucial role in our theory. This behaviour of hydrogen in GaN is qualitatively very similar to what has been observed or calculated for other materials^{1,5}, although quantitatively the values of formation energies and transition levels vary over a large range.

In some materials, however, hydrogen exhibits completely different behaviour: Fig. 1b shows that in ZnO, H^0 and H^- are always higher in energy than H^+ , for any Fermi-level position⁶. The fact that only H^+ is stable implies that hydrogen acts as a shallow donor, that is, as a source of conductivity. In this case, $\epsilon(+/-)$ does not lie in the bandgap, but above the CBM. The crucial question that we want to address here is why hydrogen can exhibit such qualitatively different behaviour. A detailed analysis showed no correlation of the

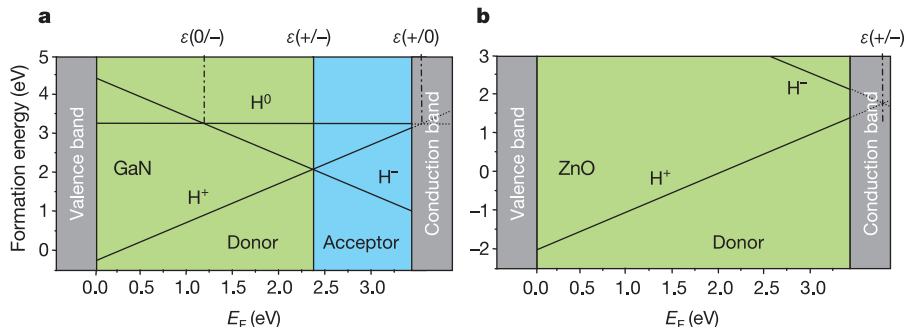


Figure 1 Formation energy of interstitial hydrogen as a function of Fermi level in two different semiconductors, illustrating the qualitatively different position of the $\epsilon(+/-)$ level. In GaN (**a**), $\epsilon(+/-)$ lies within the bandgap, whereas in ZnO (**b**), $\epsilon(+/-)$ lies above

the CBM. Energies are obtained from DFT-LDA calculations, and the range of E_F corresponds to the bandgap, with $E_F = 0$ at the VBM.

position of the charge transfer level with ionicity, bandgap, cohesive energy, or hydrogen formation energy. However, we discovered a direct link with the position of the band structure on an absolute energy scale. We illustrate this concept with the examples of GaN and ZnO. Both materials have a similar bandgap (~ 3.4 eV), but when we connect them at an interface the band edges are significantly shifted with respect to one another. This shift is expressed by the valence-band offset, ΔE_v , for which our calculations yield a value of 1.3 eV. Assuming that the hydrogen $\varepsilon(+/-)$ level occurs at the same 'absolute' energy for these two semiconductors, its position at $E_v + 2.4$ eV in GaN would place it at $E_v + 2.4 + 1.3 = E_v + 3.7$ eV in ZnO, that is, well above the ZnO CBM—consistent with explicit first-principles results for hydrogen in ZnO (ref. 6).

Figure 2 shows our results for a large number of semiconductors and insulators. The bands in Fig. 2 have been aligned using calculated valence-band offsets^{9,12–14}, which agree well with experimental data where available^{13,15}. Effects of strain or dipoles at realistic interfaces can be included using the formalism described in ref. 13, but for our present purposes we are interested in the 'natural' band line-ups between unstrained materials. Figure 2 also displays our calculated position of the $\varepsilon(+/-)$ level with respect to the VBM for selected materials. We observe that the $\varepsilon(+/-)$ values exhibit remarkable consistency in terms of their absolute energy across a range of materials. The spread in the values is less than 0.5 eV—a very narrow range, given that the band edges span a range of almost 10 eV. Note that for ZnO the $\varepsilon(+/-)$ level lies above the CBM. Also note that the alignment is quite accurate even in the case of SiO₂, an insulator with a bandgap of 9 eV.

We have carried out a number of checks to test the reliability and predictive power of this universal alignment; only a few examples are mentioned here. Figure 2 shows that the expected $\varepsilon(+/-)$ value in InN lies well above the CBM. On the basis of this prediction, explicit first-principles calculations were carried out⁸, confirming that only H⁺ is stable in InN, a result that has profound consequences for growth and doping of this technologically important compound¹⁶. Figure 2 has also correctly predicted the donor character of H in InGaAsN alloys, in which large bandgap bowing pushes the CBM down below the $\varepsilon(+/-)$ level¹⁷. Predictions are also possible for H

acting as a shallow acceptor, rather than a shallow donor. The band alignment picture indicates that $\varepsilon(+/-)$ in Ge lies close to the VBM, causing H to act as a shallow acceptor in Ge. $\varepsilon(+/-)$ lies even deeper below the VBM in GaSb, which is known to systematically exhibit unintentional p-type conductivity; the alignment picture suggests that hydrogen could be a source of this conductivity.

Figure 2 establishes a strong correlation between the location of the hydrogen $\varepsilon(+/-)$ level and the band structures on an absolute energy scale. We now address the physics behind this correlation. Alignment of impurity levels in accordance with band offsets has been previously reported, mainly for transition metals^{18,19}. Some of the arguments that were put forward to explain these trends focused on a lack of interaction between the impurity and the surrounding semiconductor, describing the electronic levels as directly linked to the vacuum level¹⁸. Such arguments are clearly inappropriate in the case of hydrogen, which forms strong bonds (with highly host-dependent bond energies) with the host atoms. Other models considered strong interaction²⁰, but only addressed alignment between semiconductors with similar lattice constants and belonging to the same class (for example, only III–V or only II–VI compounds). We do not invoke such assumptions here.

We propose the following explanation, which is grounded in explicit first-principles calculations but which captures the essence of the physics in a simple molecular-orbital description. Without losing generality, we can take a compound semiconductor CA (where C stands for cation and A for anion) as an example. In the positive charge state H forms a strong bond with the anion^{5,6,8}; this results in the creation of a dangling bond on the cation: db_C. Cation dangling bonds have energy levels near the CBM (or lowest unoccupied molecular orbital, LUMO); it is thus energetically most favourable to leave them unoccupied, amounting to the formation of db_C⁺, a cation dangling bond in the positive charge state. Conversely, in the negative charge state H forms a strong bond with the cation, leading to the creation of a dangling bond on the anion. Anion dangling bonds have energy levels near the VBM (or highest occupied molecular orbital, HOMO), and those levels prefer to be fully occupied with electrons, amounting to formation of db_A[−]. Within this picture, the formation energies of H⁺ and H[−] necessarily strongly depend on the

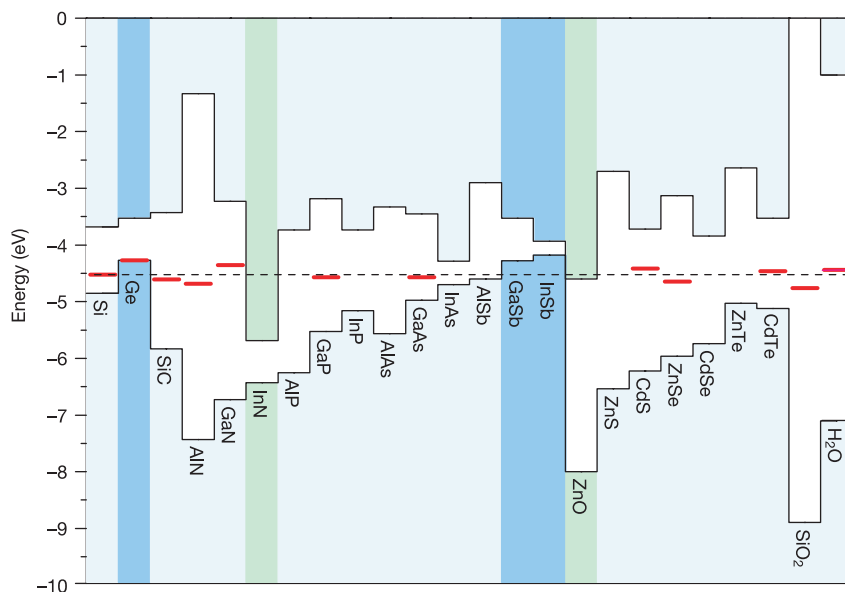


Figure 2 Band line-ups and position of the $\varepsilon(+/-)$ level for a range of semiconductors and insulators. For each material, the lower line indicates the position of the VBM, the upper line the position of the CBM, and the thick red line the position of the $\varepsilon(+/-)$ level (calculated within DFT-LDA) with respect to the VBM. The band offsets have error bars of ± 0.2 eV. We have chosen the CBM of Si to lie at a value that roughly corresponds to its electron affinity, but electron affinities in general explicitly depend on surface

conditions. For SiO₂ $\varepsilon(+/-)$ was taken from ref. 3. For H₂O, the $\varepsilon(+/-)$ level shown is the potential of the SHE (-4.44 eV), and the band edges are the calculated HOMO and LUMO levels of the H₂O molecule (in the absence of reliable information about the band structure of liquid H₂O on an absolute energy scale). The dashed line at -4.5 eV is a fit to the available data for the level positions. Materials in which H acts exclusively as a donor (acceptor) are marked in green (blue).

host (as indeed we find from first-principles calculations). But these formation energies are related in the following fashion.

Our key quantity is $\varepsilon(+/-)$, the Fermi-level position where H^+ and H^- have equal formation energy. In the model, this occurs when the Fermi level is located midway between db_C and db_A . Indeed, to form H^+ we need to remove an electron from db_C ; and to form H^- we need to place an electron in db_A . The location of E_F halfway between db_C and db_A corresponds to the 'charge neutrality' level that has been extensively discussed in the literature²¹. Indeed, it is well established that the charge neutrality level can be used to line up semiconductor band structures. If $\varepsilon(+/-)$ coincides with the charge neutrality level, then the alignment of $\varepsilon(+/-)$ in accordance with band line-ups follows directly from the established use of the charge neutrality level for this purpose. A crucial feature is that, within this model, the strengths of the individual H–C and H–A bonds (which are highly material dependent) do not enter into the determination of $\varepsilon(+/-)$. We note that, being a localized deep level, db_C is related to conduction-band states derived from the entire Brillouin zone, and not directly tied to the position of the CBM, which reflects only a single point in the zone. Similarly the db_A state is distinct from the VBM. This explains why $\varepsilon(+/-)$ can fall outside the bandgap.

We point out that the energetic position of $\varepsilon(+/-)$ on our absolute energy scale is about 4.5 eV below the vacuum level. This value is remarkably similar to an important quantity in electrochemistry, namely the chemical potential for electrons under 'standard hydrogen electrode' (SHE) conditions, which is at –4.44 eV referenced to the vacuum level (see Fig. 2). This is not a coincidence. The 'normal' or 'standard' hydrogen electrode refers to an $H^+(aq.)/H_2(g)$ electrode in water. In the framework of the present study, the relevant reaction is viewed as formation of H^+ (with H_2 as the reservoir) in an H_2O host. The equivalent of the $\varepsilon(+/-)$ level for this system can be estimated on the basis of formation energies of hydrated H_3O^+ and OH^- (refs 22, 23), resulting in a $\varepsilon(+/-)$ value between –4.2 and –4.5 eV (reflecting the spread in values for hydration enthalpies), in satisfactory agreement with the potential of the SHE. This alignment, included in Fig. 2, allows a direct connection to an important problem in electrochemistry, namely the prediction of band-edge positions of semiconductors and insulators in contact with water²⁴.

More generally, we have observed that $\varepsilon(+/-)$ levels based on proton affinities and deprotonation energies²² for a wide range of molecules fall within a narrow range. Dielectric screening, which is implicitly present in a liquid environment and plays a central role in the charge neutrality level²¹, narrows this range even further. The fact that the universality of the alignment extends to aqueous solutions highlights the commonality of driving forces for hydrogen chemistry across a wide range of systems. □

Methods

The calculations for interstitial H were performed in supercells containing either 64 atoms for zinc blende or 96 atoms for wurtzite. Effects of semicore d states (such as the $3d$ states in Ga or Zn) were included using the nonlinear core correction¹¹. For H the Coulomb potential is used. Energy differences for hydrogen-containing systems are well converged at a 40-Ry plane-wave cut-off. For each charge state of H, many possible sites in the lattice were explored and the global minimum was identified. We estimate the error bar on $\varepsilon(+/-)$ due to DFT-LDA errors to be ~ 0.4 eV, based on comparisons between calculations with different pseudopotentials (particularly those that result in different bandgaps owing to different treatments of the semicore d states). The bandgap error inherent in DFT-LDA has only minor effects on $\varepsilon(+/-)$, because the formation energies of H^+ and H^- are not sensitive to bandgap corrections: for H^+ , the defect-induced state is unoccupied, whereas for H^- , the state is related to an anion dangling bond and thus valence-band derived. Calculations for band line-ups were performed for nonpolar interfaces using the methodology of ref. 12. The line-up for Si/SiO₂ was taken from ref. 25.

Received 29 November 2002; accepted 24 April 2003; doi:10.1038/nature01665.

1. Pankove, J. I. & Johnson, N. M. *Semiconductors and Semimetals* Vol. 34, *Hydrogen in Semiconductors* (Boston, Academic, 1991).
2. Blöchl, P. & Stathis, J. H. Hydrogen electrochemistry and stress-induced leakage current in silica. *Phys. Rev. Lett.* **62**, 372–375 (1999).
3. Schlappbach, L. & Züttel, A. Hydrogen-storage materials for mobile applications. *Nature* **414**, 353–358 (2001).

4. Mills, B. E., Martin, R. L. & Shirley, D. A. Further studies of the core binding energy-proton affinity correlation in molecules. *J. Am. Chem. Soc.* **98**, 2380–2385 (1976).
5. Van de Walle, C. G. & Johnson, N. M. in *Semiconductors and Semimetals* Vol. 57, *Gallium Nitride (GaN) II* (eds Pankove, J. I. & Moustakas, T. D.) 157–184 (Academic, Boston, 1998).
6. Van de Walle, C. G. Hydrogen as a cause of doping in ZnO. *Phys. Rev. Lett.* **85**, 1012–1015 (2000).
7. Hofmann, D. M. *et al.* Hydrogen: a relevant shallow donor in zinc oxide. *Phys. Rev. Lett.* **88**, 045504 (2002).
8. Limpijumpong, S. & Van de Walle, C. G. Passivation and doping due to hydrogen in III-nitrides. *Phys. Status Solidi B* **228**, 303–307 (2001).
9. Franciosi, A. & Van de Walle, C. G. Heterojunction band offset engineering. *Surf. Sci. Rep.* **25**, 1–140 (1996).
10. Bockstedte, M., Kley, A., Neugebauer, J. & Scheffler, M. Density-functional theory calculations for poly-atomic systems: Electronic structure, static and elastic properties and *ab initio* molecular dynamics. *Comput. Phys. Commun.* **107**, 187–222 (1997).
11. Louie, S. G., Froyen, S. & Cohen, M. L. Nonlinear ionic pseudopotentials in spin-density-functional calculations. *Phys. Rev. B* **26**, 1738–1742 (1982).
12. Van de Walle, C. G. & Martin, R. M. Theoretical calculations of heterojunction discontinuities in the Si/Ge system. *Phys. Rev. B* **34**, 5621–5634 (1986).
13. Van de Walle, C. G. Band lineups and deformation potentials in the model-solid theory. *Phys. Rev. B* **39**, 1871–1883 (1989).
14. Majewski, J. A., Städele, M. & Vogl, P. Stability and band offsets of SiC/GaN, SiC/AlN, and AlN/GaN heterostructures. *Mater. Res. Soc. Symp. Proc.* **449**, 917–922 (1997).
15. Yu, E. T., McCaldin, J. O. & McGill, T. C. in *Solid State Physics* Vol. 46 (eds Ehrenreich, H. & Turnbull, D.) 1–146 (Academic, Boston, 1992).
16. Look, D. C. *et al.* Donor and acceptor concentrations in degenerate InN. *Appl. Phys. Lett.* **80**, 258–260 (2002).
17. Janotti, A., Zhang, S. B., Wei, S.-H. & Van de Walle, C. G. The effects of hydrogen on the electronic properties of GaAsN alloys. *Phys. Rev. Lett.* **89**, 086403 (2002).
18. Ledebor, L. Å. & Ridley, B. K. On the position of energy levels related to transition-metal impurities in III-V semiconductors. *J. Phys. C* **15**, L961–L964 (1982).
19. Caldas, M. J., Fazzio, A. & Zunger, A. A universal trend in the binding energies of deep impurities in semiconductors. *Appl. Phys. Lett.* **45**, 671–673 (1984).
20. Tersoff, J. & Harrison, W. A. Transition-metal impurities in semiconductors—their connection with band lineups and Schottky barriers. *Phys. Rev. Lett.* **58**, 2367–2370 (1987).
21. Tersoff, J. Theory of semiconductor heterojunctions: The role of quantum dipoles. *Phys. Rev. B* **30**, 4874–4877 (1984).
22. NIST Chemistry WebBook (<http://webbook.nist.gov/chemistry>) (March 2003).
23. Mejias, J. A. & Lago, S. Calculation of the absolute hydration enthalpy and free energy of H^+ and OH^- . *J. Chem. Phys.* **113**, 7306–7316 (2000).
24. Grätzel, M. Photoelectrochemical cells. *Nature* **414**, 338–344 (2001).
25. Tuttle, B. R. *Ab initio* valence band offsets between Si(100) and SiO₂ from microscopic models. *Phys. Rev. B* **67**, 155324 (2003).

Acknowledgements We thank S. Limpijumpong, M. Fuchs, M. Chabiny, D. Biegelsen and the late J. McCaldin for discussions and support. This work was supported in part by the Air Force Office of Scientific Research and by the Deutsche Forschungsgemeinschaft.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.G.V.d.W. (vandewalle@parc.com).

Chaperonin-mediated stabilization and ATP-triggered release of semiconductor nanoparticles

Daisuke Ishii*, Kazushi Kinbara*, Yasuhiro Ishida*, Noriyuki Ishii†, Mina Okochi‡, Masafumi Yohda‡ & Takuzo Aida*

* Department of Chemistry and Biotechnology, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

† Biological Information Research Center, National Institute of Advanced Industrial Science and Technology, Tsukuba Central 6, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan

‡ Department of Biotechnology, Faculty of Technology, Tokyo University of Agriculture and Technology, 2-24-16 Naka-cho, Koganei, Tokyo 184-8588, Japan

Various properties of semiconductor nanoparticles, including photoluminescence and catalytic activity, make these materials attractive for a range of applications^{1,2}. As nanoparticles readily coagulate and so lose their size-dependent properties, shape-persistent three-dimensional stabilizers that enfold nanoparti-

cles have been exploited^{3–9}. However, such wrapping approaches also make the nanoparticles insensitive to external stimuli, and so may limit their application. The chaperonin proteins GroEL (from *Escherichia coli*) and *T.th* (*T.th* cpn, from *Thermus thermophilus* HB8) encapsulate denatured proteins inside a cylindrical cavity; after refolding, the encapsulated proteins are released by the action of ATP inducing a conformational change of the cavity^{10,11}. Here we report that GroEL and *T.th* cpn can also enfold CdS semiconductor nanoparticles, giving them high thermal and chemical stability in aqueous media. Analogous to the biological function of the chaperonins, the nanoparticles can be readily released from the protein cavities by the action of ATP. We expect that integration of such biological mechanisms into materials science will open a door to conceptually new bio-responsive devices.

The chaperonin GroEL consists of two supramolecular rings that are stacked to form a double-decker architecture (Fig. 1a)¹². (Each ring consists of seven protein subunits; each subunit has a molecular mass of 60 kDa.) GroEL possesses a cylindrical cavity with a diameter of 4.5 nm, and a wall thickness of 4.6 nm (ref. 13). The chaperonin *T.th* cpn originates from the thermophilic bacterium, *Thermus thermophilus* HB8^{14,15}, and consists of a GroEL-like tetradecamer (840 kDa) of a protein subunit called cpn60 (60 kDa), which is hybridized, with a capping protein assembly [cpn10]₇ (70 kDa) on either side of its cylindrical cavity (Fig. 1a)^{16,17}. Transmission electron microscopy (TEM) showed that the cavity size and wall thickness of *T.th* cpn (~5 nm) are both comparable to those of GroEL.

We prepared CdS nanoparticles (2–4 nm) according to a method reported in ref. 18. For the complexation with chaperonins, a dimethylformamide (DMF) solution of CdS nanoparticles was added to a Tris–HCl buffer solution of GroEL or *T.th* cpn. In the presence of the chaperonin proteins, the photoluminescence of CdS nanoparticles at 430–720 nm (excitation wavelength $\lambda_{\text{ext}} = 370$ nm) lasted for an unusually long period of time without decay (for example, 400 days with GroEL). In the absence of the chaperonins, the characteristic fluorescence disappeared within 2 hours.

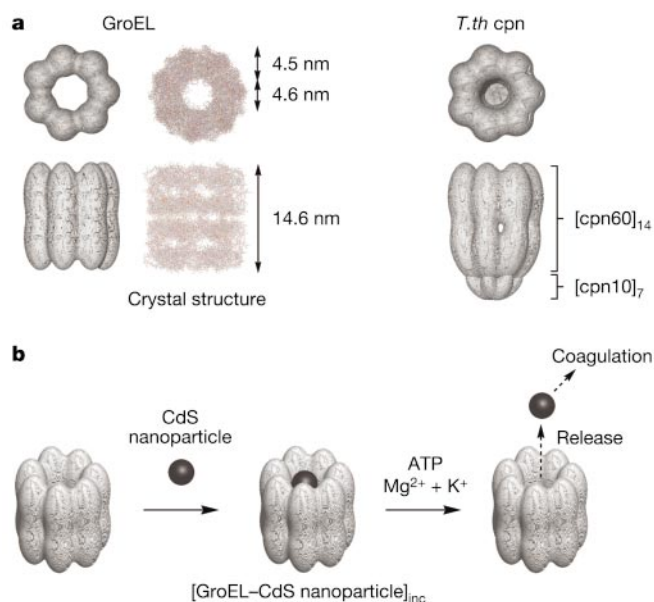


Figure 1 Chaperonin proteins as ATP-responsive barrels for inclusion of nanoparticles. **a**, Schematic illustrations (top and side views) of GroEL (crystal structures, right)¹³ and *T.th* cpn. **b**, Schematic representation of the formation of GroEL–CdS nanoparticle complexes by inclusion of CdS nanoparticles into the cylindrical cavity of GroEL, and its ATP-triggered guest release.

The tubular structure of the chaperonin proteins was essential for the stabilization of CdS nanoparticles. Fluorescence was not enhanced by the use of GroES, a bowl-shaped, [cpn10]₇-like GroEL-capping protein, in place of GroEL.

We isolated complexes of *T.th* cpn or GroEL with CdS nanoparticles by size-exclusion chromatography (SEC). By means of inductively coupled plasma mass spectrometry (ICP-MS), we confirmed that the complexes both contained Cd²⁺. The *T.th* cpn–nanoparticle complex showed circular dichroism bands, whose intensities were identical to those of intact *T.th* cpn. An analytical SEC trace of this solution with an ultraviolet (UV)/fluorescence dual detector showed single, sharp elution peaks, which were superimposable with one another at nearly the same elution volume as intact *T.th* cpn (Fig. 2a). As intact *T.th* cpn is barely fluorescent, the above results strongly indicate that the CdS nanoparticles co-localized with *T.th* cpn to form an inclusion complex. The photoluminescence of the complexes at 430–720 nm upon excitation at 370 nm is characteristic of CdS nanoparticles (Supplementary Fig. S1)^{19,20}. We also conducted analytical SEC of the *T.th* cpn–CdS inclusion complexes using a refractive index (RI)/multi-angle light scattering dual detector, where the molecular mass of the complex was evaluated to be 915 ± 4 kDa, which is nearly identical to that of intact *T.th* cpn (910 kDa) (Supplementary Fig. S2). These results demonstrate that *T.th* cpn within the protein–nanoparticle complex preserves its own structural identity, without formation of higher aggregates or dissociation into the protein subunits.

We also succeeded in imaging these complexes by TEM. Figure 3a shows a TEM picture of the *T.th* cpn–CdS inclusion complex, isolated by SEC and stained with uranyl acetate, where the bullet-like side view and the doughnut-like end-on view of *T.th* cpn protein are clearly observed, as reported previously^{16,17}. The TEM picture of the *T.th* cpn–CdS inclusion complex in the end-on orientation shows a darker central region when compared to intact *T.th* cpn (Fig. 3c, Supplementary Fig. S4), owing to the presence of a CdS nanoparticle within the protein cavity (Fig. 3b). No CdS nanoparticles were observed on the exterior of the protein surface. When a single inclusion complex was irradiated for 1 minute by a focused electron beam, the protein was partially destroyed, so that the CdS nanoparticle included within the chaperonin cavity was more clearly imaged by TEM (Supplementary Fig. S3). From the TEM image under such partially destructive conditions, only a single nanoparticle appeared to be present within the chaperonin cavity. As the CdS nanoparticles are positively charged owing to Cd²⁺ ions on the surface, it is reasonable to conclude that a second nanoparticle will not be included in the chaperonin cavity owing to electrostatic repulsion. For statistical analysis of the TEM picture, we randomly selected 793 images of end-on *T.th* cpn–CdS nanoparticle complexes, and found that 75% showed a dark central cavity due to the presence of a CdS nanoparticle.

The CdS nanoparticles within the chaperonin proteins are electronically insulated. The photoexcited singlet state of CdS nanoparticles is known to be quenched by an electron acceptor such as methyl viologen (MV²⁺) via photoinduced electron transfer from the former to the latter^{21,22}. Fluorescence titration of CdS nanoparticles in DMF with MV²⁺ at 25 °C gave Stern–Volmer plots (Fig. 4), from which the Stern–Volmer constant was evaluated to be 3.3. In sharp contrast, the fluorescence quenching of the inclusion complex with GroEL took place much less efficiently even at a high concentration of MV²⁺, where the observed Stern–Volmer constant (0.20) was more than 15 times smaller than that for intact CdS nanoparticles. We also found an even smaller Stern–Volmer constant for the *T.th* cpn–nanoparticle complex (0.09). These observations are quite reasonable, considering that the thickness of the protein envelope around the CdS nanoparticle (4.6 nm for GroEL¹³) exceeds an upper limit of the electron-transferable distance (~3 nm) in ordinary systems.

Chaperonins belong to the family of heat-shock proteins²³, and

T.th cpn, originating from a thermophilic bacterium, has been reported to be stable against thermal denaturation up to 80 °C (ref. 14). We found that the *T.th* cpn–CdS nanoparticle complex is also thermally stable, and maintains its characteristic photoluminescence activity up to 80 °C. In heating–cooling cycles in a temperature range of 4–80 °C (Fig. 5a), the photoluminescence of the inclusion complex in Tris–HCl buffer was quenched upon heating, but recovered the original intensity upon cooling down to 4 °C. However, when heated at 90 °C for 10 min, the protein component underwent partial denaturation as observed by circular dichroism, and the complex irreversibly lost 18% of its photoluminescence activity (Fig. 5a, arrow). Although intact CdS nanoparticles in DMF also showed a similar temperature-dependent photoluminescence profile, the response became irreversible upon heating at only 30 °C, forming a colloidal precipitate of CdS. A GroEL–CdS nanoparticle complex showed thermoresponsive photoluminescence activity (Fig. 5b), but became partially irreversible upon heating for 20 min at 60 °C (Fig. 5b, arrow), just below the denaturation temperature of intact GroEL²⁴. Thus, the thermal

stability of the included nanoparticle is totally dependent on that of the chaperonin protein.

Chaperonin proteins, upon binding with ATP in the presence of Mg^{2+} and K^{+} , undergo a large conformational change, which results in the release of refolded guest proteins from the cavity^{25,26}. We found that *T.th* cpn–CdS nanoparticle complexes also respond to ATP, and release included nanoparticles from the cavities under similar conditions (Fig. 1b). When a Tris–HCl buffer solution of ATP containing $MgCl_2$ was added to a buffer solution of *T.th* cpn–CdS nanoparticle complexes containing KCl, the mixture turned slightly cloudy within seconds to give colloidal substances, where the supernatant solution after centrifugation was no longer fluorescent (Fig. 2c). The release of CdS nanoparticles from *T.th* cpn complexes by the action of ATP was clearly demonstrated by analytical SEC with a UV/fluorescence dual detector (Fig. 2b). After the addition of ATP, the UV response of the SEC trace of the complexes showed a sharp elution peak assignable to *T.th* cpn and an additional broad peak in the lower-molecular-mass region due to ATP and its hydrolysed products, while no fluorescence responses

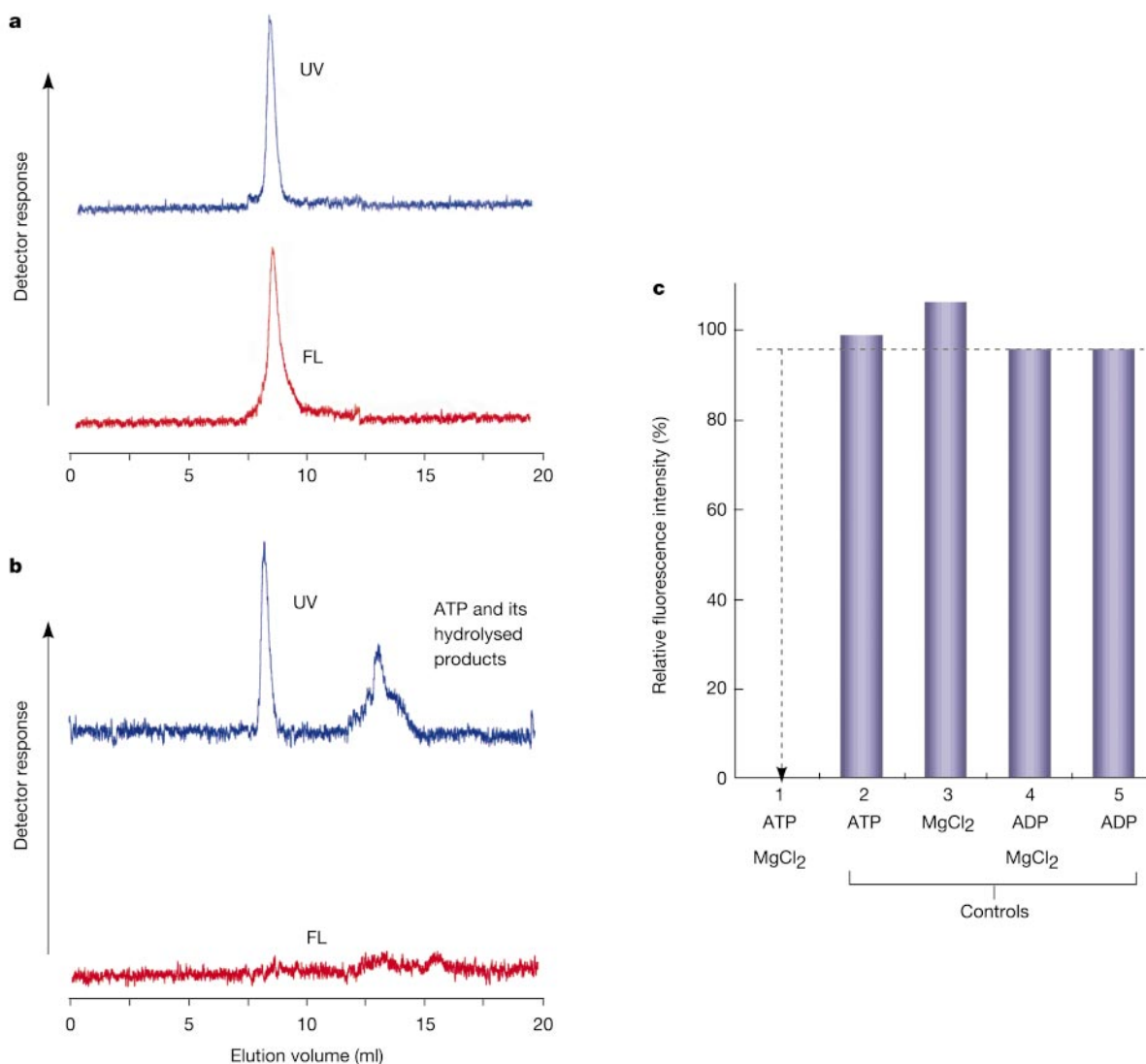


Figure 2 SEC analysis of *T.th* cpn–CdS the nanoparticle inclusion complex and its specific response to ATP in KCl-containing Tris–HCl buffer. **a**, Analytical SEC trace of *T.th* cpn–CdS inclusion complexes monitored by an UV (upper chart)/fluorescence (FL) (lower chart) dual detector, and **b**, that of *T.th* cpn–CdS inclusion complexes after addition of ATP and

$MgCl_2$. The peak at 12.5–15.0 ml is derived from ATP and its hydrolysed products. **c**, Changes in fluorescence intensity of *T.th* cpn–CdS complexes in Tris–HCl buffer: 1, ATP + $MgCl_2$; 2, ATP; 3, $MgCl_2$; 4, ADP + $MgCl_2$; 5, ADP.

were observed for these two peaks. The fraction corresponding to *T.th* cpn, isolated by SEC from the reaction mixture, lost most (nearly 90%) of its Cd^{2+} , as determined by ICP-MS. On the other hand, addition of ATP or MgCl_2 alone to the KCl-containing buffer solution of the inclusion complexes neither resulted in the generation of colloidal substances nor fluorescence quenching (Fig. 2c)—in contrast, a DMF solution of intact CdS nanoparticles immediately lost its luminescence activity upon addition of MgCl_2 .

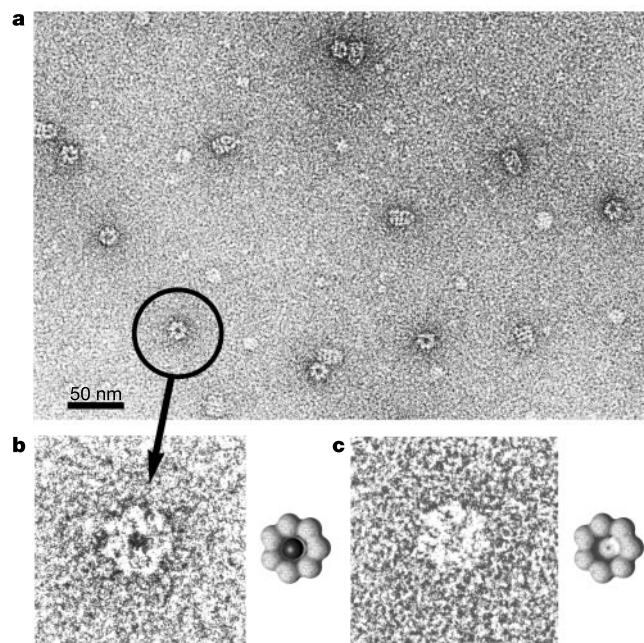


Figure 3 Transmission electron micrographs of *T.th* cpn–CdS nanoparticle complexes and intact *T.th* cpn. **a**, **b**, *T.th* cpn–CdS nanoparticle complexes isolated by SEC. **c**, Intact *T.th* cpn. The models on the right sides of the images in **b** and **c** are schematic representations of the end-view of the complex and intact *T.th* cpn, respectively.

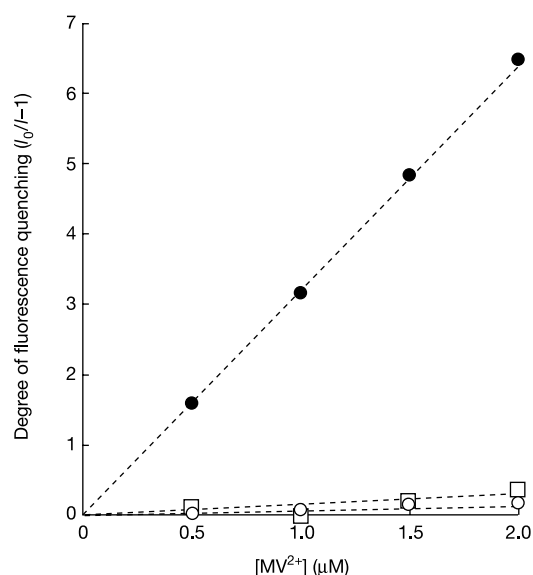


Figure 4 Shielding effects of chaperonins on electron transfer from CdS nanoparticles to methyl viologen (MV^{2+}). Stern–Volmer plots for fluorescence titration with MV^{2+} at 25 °C of a DMF solution of CdS nanoparticle (filled circles), a Tris–HCl buffer solution containing GroEL–CdS nanoparticle inclusion complexes (open squares), and a Tris–HCl buffer solution of *T.th* cpn–CdS nanoparticle inclusion complexes (open circles). I and I_0 represent the intensities of fluorescence and its initial value, respectively.

Thus, *T.th* cpn makes the included CdS nanoparticle thermally stable, tolerant to electrolytes, and electronically dormant, but can readily release the guest nanoparticle from its cylindrical cavity upon binding with ATP. As expected, the *T.th* cpn–CdS nanoparticle complexes hardly responded to ADP (adenosine-5'-diphosphate)—there was no substantial change in the fluorescence intensity, regardless of the presence or absence of Mg^{2+} (Fig. 2c). We also found that GroEL–CdS nanoparticle complexes behave similarly upon binding with ATP. □

Methods

Sample preparation

To a N_2 -purged DMF solution of $\text{Cd}(\text{OAc})_2 \cdot 2\text{H}_2\text{O}$ (30 ml, 5.56 mM), with vigorous stirring, was added dropwise a MeOH solution of $\text{Na}_2\text{S} \cdot 9\text{H}_2\text{O}$ (3 ml, 5.56 mM) at -40°C . The resulting mixture was stirred for 1.5 h to afford a yellow solution, which showed absorption and emission spectra ($\lambda_{\text{ext}} = 370$ nm, emission wavelength $\lambda_{\text{em}} = 530$ nm) characteristic of nanosized CdS particles. TEM analysis showed that the resulting CdS nanoparticles have a size distribution in which the majority were in the diameter range 2–4 nm. To a 1.9-ml Tris–HCl buffer (25 mM, pH 7.5 with 100 mM KCl) solution of *T.th* cpn (0.5 μM) was added dropwise a 100- μl DMF solution of CdS nanoparticles (5.56 mM based on Cd^{2+}) at 4 °C. The mixture was allowed to stand at 4 °C until the initial spectral change subsided (~ 100 h), and then subjected to SEC using Tris–HCl buffer (25 mM, pH 7.5 with 100 mM KCl) as eluent on Sephacryl S-300 HR or S-400 HR with an ÄKTA-prime low-pressure chromatography system (Amersham Pharmacia Biotech). SEC traces at 280 nm allowed the isolation of a protein-containing fraction, which was concentrated by ultrafiltration with an USY-5 ultrafilter unit (Advantec) at 4 °C, affording 2 ml of a Tris–HCl buffer solution of the *T.th* cpn–CdS nanoparticle complex (0.1 μM). Preparation and isolation of GroEL–CdS nanoparticle inclusion complexes were performed in a similar fashion.

Characterization

Analytical SEC traces (TSK gel G4000SW_{XL} (Tosoh); eluent, Tris–HCl buffer 25 mM, pH 7.5 with 100 mM KCl) of *T.th* cpn–CdS nanoparticle inclusion complexes were followed by UV (observed at 280 nm) and fluorescence (excitation at 370 nm, observed at 530 nm) responses by UV970 and FP2020-plus spectrophotometers (JASCO), respectively, at a flow rate of 0.5 ml min^{−1}. Analytical SEC (Tosoh TSK gel G4000SW_{XL} or Shodex Protein KW-804) of the complexes was performed using an UV/fluorescence dual detector and an RI/multi-angle light scattering dual detector, at a flow rate of 0.5 ml min^{−1}.

TEM imaging

Samples were applied to an electron microscope specimen grid covered with a thin carbon support film that had been hydrophilized by ion bombardment. After drying with a pre-water-soaked filter paper, the grid was negatively stained with 1% uranyl acetate. TEM micrographs were then recorded on a JEM-2010 TEM (JEOL) operating at an anode voltage of 120 kV.

Methyl viologen titration

To a DMF solution of CdS nanoparticles (2 ml, 700 nM based on Cd^{2+}) were added 10- μl aliquots of a 100- μM Tris–HCl buffer (25 mM, pH 7.5 with 100 mM KCl) solution of

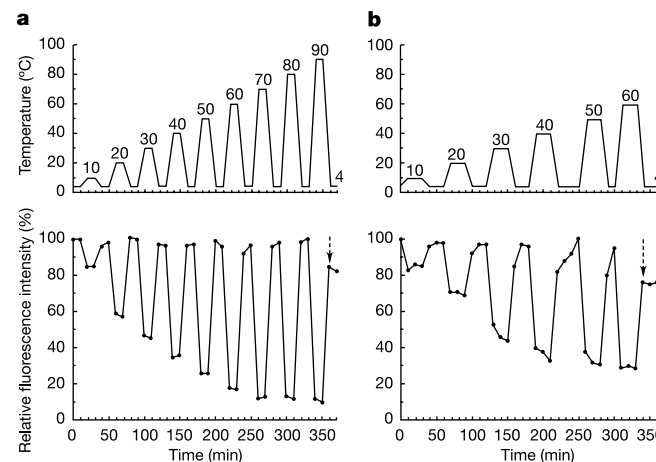


Figure 5 Thermal stabilities and relative fluorescence intensities of GroEL–CdS and *T.th* cpn–CdS inclusion complexes. Top, temperature-dependent photoluminescence profiles of **a**, *T.th* cpn–CdS and **b**, GroEL–CdS nanoparticle inclusion complexes in Tris–HCl buffer. Fluorescence intensities at 4 °C were used as the bases for relative fluorescence intensities (bottom). Numbers of plateau in top panels denote annealing temperatures.

MV²⁺. Likewise, Tris–HCl buffer (25 mM, pH 7.5 with 100 mM KCl) solutions of GroEL–CdS complex and *T.th* cpn–CdS complex (2 ml, 700 nm based on Cd²⁺) were titrated. Emission ($\lambda_{\text{ext}} = 370$ nm) spectra were recorded on a FP-777W spectrophotometer (JASCO).

Thermal stability

Fluorescence spectra ($\lambda_{\text{ext}} = 370$ nm, wavelength of observed fluorescence $\lambda_{\text{obsd}} = 530$ nm) of GroEL–CdS complexes and *T.th* cpn–CdS complexes were recorded at designated temperatures on a FP-777W spectrophotometer (JASCO), where the fluorescence intensities at 4 °C were used as the bases for relative fluorescence intensities. The temperature was directly controlled by a ECT271 Peltier thermometric apparatus (JASCO; 40 °C min^{−1} on heating and 25 °C min^{−1} on cooling).

ATP response

To a 2-ml Tris–HCl buffer (25 mM, pH 7.5 with 100 mM KCl) solution of *T.th* cpn–CdS complexes (0.5 μ M based on *T.th* cpn) were added aqueous solutions of ATP (100 mM) and MgCl₂ (1 M) ([ATP] = 20 μ M, [Mg²⁺] = 25 mM after mixing), and the mixture was incubated at 70 °C for 10 min. The supernatant solution was subjected to fluorescence spectroscopy and analytical SEC with an UV/fluorescence dual detector.

Received 11 November 2002; accepted 14 April 2003; doi:10.1038/nature01663.

1. Alivisatos, A. P. Semiconductor clusters, nanocrystals, and quantum dots. *Science* **271**, 933–937 (1996).
2. Schmid, G. et al. Current and future applications of nanoclusters. *Chem. Soc. Rev.* **28**, 179–185 (1999).
3. Meldrum, F. C., Heywood, B. R. & Mann, S. Magnetoferritin — *in vitro* synthesis of a novel magnetic protein. *Science* **257**, 522–523 (1992).
4. Wong, K. K. W. & Mann, S. Biomimetic synthesis of cadmium sulfide-ferritin nanocomposites. *Adv. Mater.* **8**, 928–932 (1996).
5. Shenton, W., Pum, D., Sleytr, U. B. & Mann, S. Biocrystal templating of CdS superlattices using self-assembled bacterial S-layers. *Nature* **389**, 585–587 (1997).
6. Balogh, L. & Tomalia, D. A. Poly(amidoamine) dendrimer-templated nanocomposites. I. Synthesis of zerovalent copper nanoclusters. *J. Am. Chem. Soc.* **120**, 7355–7356 (1998).
7. Lemon, B. I. & Crooks, R. M. Preparation and characterization of dendrimer-encapsulated CdS semiconductor quantum dots. *J. Am. Chem. Soc.* **122**, 12886–12887 (2000).
8. Shenton, W., Mann, S., Colfen, H., Bacher, A. & Fischer, M. Synthesis of nanophasic iron oxide in lumazine synthase capsids. *Angew. Chem. Int. Edn Engl.* **40**, 442–445 (2001).
9. McMillan, R. A. et al. Ordered nanoparticle arrays formed on engineered chaperonin protein templates. *Nature Mater.* **1**, 247–252 (2002).
10. Roseman, A. M., Chen, S., White, H., Braig, K. & Saibil, H. R. The chaperonin ATPase cycle: mechanism of allosteric switching and movements of substrate-binding domains in groEL. *Cell* **87**, 241–251 (1996).
11. Ranson, N. A. et al. ATP-bound states of groEL captured by cryo-electron microscopy. *Cell* **107**, 869–879 (2001).
12. Hendrix, R. W. Purification and properties of groE, a host protein involved in bacteriophage assembly. *J. Mol. Biol.* **129**, 375–392 (1979).
13. Braig, K. et al. The crystal structure of the bacterial chaperonin groEL at 2.8 Å. *Nature* **371**, 578–586 (1994).
14. Taguchi, H., Konishi, J., Ishii, N. & Yoshida, M. A chaperonin from a thermophilic bacterium, *Thermus thermophilus*, that controls refoldings of several thermophilic enzymes. *J. Biol. Chem.* **266**, 22411–22418 (1991).
15. Amada, K. et al. Molecular cloning, expression, and characterization of chaperonin-60 and chaperonin-10 from a thermophilic bacterium, *Thermus thermophilus* HB8. *J. Biochem.* **118**, 347–354 (1995).
16. Ishii, N., Taguchi, H., Sumi, M. & Yoshida, M. Structure of holo-chaperonin studied with electron microscopy. Oligomeric cpn10 on top of two layers of cpn60 rings with two stripes each. *FEBS Lett.* **299**, 169–174 (1992).
17. Ishii, N., Taguchi, H., Sasabe, H. & Yoshida, M. Folding intermediate binds to the bottom of bullet-shaped holo-chaperonin and is readily accessible to antibody. *J. Mol. Biol.* **236**, 691–696 (1994).
18. Murakoshi, K. et al. Preparation of size-controlled hexagonal CdS nanocrystallites and the characteristics of their surface structures. *J. Chem. Soc. Faraday Trans.* **94**, 579–586 (1998).
19. Hosokawa, H. et al. *In-situ* EXAFS observation of the surface structure of colloidal CdS nanocrystallites in *N,N*-dimethylformamide. *J. Phys. Chem.* **100**, 6649–6656 (1996).
20. Brus, L. E. Electron-electron and electron-hole interactions in small semiconductor crystallites: The size dependence of the lowest excited electronic state. *J. Chem. Phys.* **80**, 4403–4409 (1984).
21. Ramsden, J. J. & Grätzel, M. Photoluminescence of small cadmium sulfide particles. *J. Chem. Soc. Faraday Trans. 1* **80**, 919–933 (1984).
22. Henglein, A. Photochemistry of colloidal cadmium sulfide. 2. Effects of adsorbed methyl viologen and of colloidal platinum. *J. Phys. Chem.* **86**, 2291–2293 (1982).
23. Llorca, O., Galán, A., Carrascosa, J. L., Muga, A. & Valpuesta, J. M. GroEL under heat-shock. *J. Biol. Chem.* **273**, 32587–32594 (1998).
24. Martin, J., Horwich, A. L. & Hartl, F. U. Prevention of protein denaturation under heat stress by the chaperonin Hsp60. *Science* **258**, 995–998 (1992).
25. Goloubinoff, P., Christeller, J. T., Gatenby, A. A. & Lorimer, G. H. Reconstitution of active dimeric ribulose biphosphate carboxylase from an unfolded state depends on two chaperonin proteins and Mg-ATP. *Nature* **342**, 884–889 (1989).
26. Todd, M. J., Viitanen, P. V. & Lorimer, G. H. Hydrolysis of adenosine 5'-triphosphate by *Escherichia coli* groEL: effects of groES and potassium ion. *Biochemistry* **32**, 8560–8567 (1993).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank K. Konishi for his initial contribution to the present work; K. Tsumoto for discussions; J. Oono and M. Nakamura for SEC analysis with MALS. N.I. was responsible for TEM microscopy. We acknowledge support from the 21st Century COE Programs of Research and Education (T.A., Human-Friendly Materials Based on Chemistry; M.Y., Future Nano-Materials), and from the JST ERATO Nanospace program. K.K. acknowledges support from the Nissan Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.A. (aida@macro.t.u-tokyo.ac.jp).

Evidence for low sulphate and anoxia in a mid-Proterozoic marine basin

Yanan Shen*, Andrew H. Knoll* & Malcolm R. Walter†

* Botanical Museum, Harvard University, 26 Oxford Street, Cambridge, Massachusetts 02138, USA

† Australian Centre for Astrobiology, Macquarie University, New South Wales 2109, Australia

Many independent lines of evidence document a large increase in the Earth's surface oxidation state 2,400 to 2,200 million years ago^{1–4}, and a second biospheric oxygenation 800 to 580 million years ago, just before large animals appear in the fossil record^{5,6}. Such a two-staged oxidation implies a unique ocean chemistry for much of the Proterozoic eon, which would have been neither completely anoxic and iron-rich as hypothesized for Archaean seas, nor fully oxic as supposed for most of the Phanerozoic eon⁷. The redox chemistry of Proterozoic oceans has important implications for evolution⁸, but empirical constraints on competing environmental models are scarce. Here we present an analysis of the iron chemistry of shales deposited in the marine Roper Basin, Australia, between about 1,500 and 1,400 million years ago, which record deep-water anoxia beneath oxidized surface water. The sulphur isotopic compositions of pyrites in the shales show strong variations along a palaeodepth gradient, indicating low sulphate concentrations in mid-Proterozoic oceans. Our data help to integrate a growing body of evidence favouring a long-lived intermediate state of the oceans, generated by the early Proterozoic oxygen revolution and terminated by the environmental transformation late in the Proterozoic eon.

The Roper Group is a thick (>1,500 m), predominately marine siliciclastic succession preserved over an area of 145,000 km² west of the Gulf of Carpenteria in Australia's Northern Territory. The ramp-like succession, formed in an intracratonic basin in contact with the global ocean, contains six major depositional sequences that record episodic flooding followed by basin filling and shoreline progradation⁹. Facies are laterally persistent and repeat vertically in successive sequences, allowing environmental variations in seawater chemistry to be distinguished from long-term secular changes. Zircons from a tuff low in the succession have yielded a U–Pb sensitive high-resolution ion microprobe age of 1,492 ± 3 Myr⁹. Our samples were collected from four drill cores (Urapunga 4, 5; Amoco 82/3; Golden Grove 1) (Fig. 1): there were a total of 117 samples—37 from inner-shelf shales, 37 from distal-shelf facies, and 43 from basinal environments, as interpreted⁹ using sedimentological criteria.

To investigate the redox chemistry of the Roper seaway, we examined Fe species—including dithionite-extractable Fe (Fe_D), pyrite Fe (Fe_P), HCl-extractable Fe (Fe_H), and total Fe (Fe_T)—in carbonaceous shales. Two ratios, that between highly reactive Fe

($\text{Fe}_{\text{HR}} = \text{Fe}_{\text{D}} + \text{Fe}_{\text{P}}$) and Fe_{T} , and the ratio $\text{Fe}_{\text{P}}/(\text{Fe}_{\text{P}} + \text{Fe}_{\text{H}})$, known as the degree of pyritization (DOP), have been used successfully to evaluate the redox state of ancient oceans^{10–13}. $\text{DOP} < 0.45$ and $\text{Fe}_{\text{HR}}/\text{Fe}_{\text{T}} < 0.38$ are generally found for sediments depositing from oxic bottom water, whereas $\text{DOP} > 0.45$ and $\text{Fe}_{\text{HR}}/\text{Fe}_{\text{T}} > 0.38$ may be diagnostic for euxinic sediments^{10,12}. The enrichment of Fe_{HR} and high DOP in euxinic sediments arises fundamentally from pyrite formation in the upper zone of the sulphidic water column, where dissolved iron concentrations can be high^{10,13}. The extra reactive iron in euxinic water columns originates primarily from the reduction of iron oxides in basin margin sediments that can be transported to the ocean interior, or from iron oxide containing particles falling through the water column^{12–13}. In contrast, when bottom waters are oxygenated, pyrite can only form where anoxic conditions become established within the sediment¹⁰. Consequently, because water-column Fe scavenging is absent, oxic marine sediments are characterized by low $\text{Fe}_{\text{HR}}/\text{Fe}_{\text{T}}$ and DOP values^{10,12}.

Roper shales from inner-shelf environments have low $\text{Fe}_{\text{HR}}/\text{Fe}_{\text{T}}$ (0.03–0.18; mean 0.07) and low DOP (0.00–0.34; mean 0.12). Similarly, Fig. 1 shows that $\text{Fe}_{\text{HR}}/\text{Fe}_{\text{T}}$ and DOP for most distal-shelf shales fall within the ‘oxic’ quadrant defined by $\text{Fe}_{\text{HR}}/\text{Fe}_{\text{T}} < 0.38$ and $\text{DOP} < 0.45$ (full analytical data are available as Supplementary Information). In contrast, basal sediments are dominated by high $\text{Fe}_{\text{HR}}/\text{Fe}_{\text{T}}$ (>0.38) and DOP (>0.45) values, indicating euxinic bottom waters (Fig. 1; see also ref. 14). The relatively low $\text{Fe}_{\text{HR}}/\text{Fe}_{\text{T}}$ and DOP in some basal shales of the Roper Group could result from the decreased transfer of dissolved Fe from shelf sediments or from a high flux of terrestrial clastics^{12,13}. Alternatively, it may be that anoxia was established at depths greater than the shelf–basin boundary (defined in ref. 9 on the basis of physical sedimentary features). Regardless, high DOP and $\text{Fe}_{\text{HR}}/\text{Fe}_{\text{T}}$ together provide compelling evidence for the presence of predominantly euxinic deep waters beneath oxic surface waters at the time that Roper sediments were deposited.

We measured the S isotopic compositions of sedimentary pyrites along the same palaeoenvironmental gradient from inner-shelf to deep basal facies. Although the literature is replete with data showing stratigraphic variation in the S isotopic compositions of sedimentary pyrite, the data presented here are, to the best of our knowledge, the first to be integrated fully with the sequence stratigraphy of a Proterozoic basin, allowing us to understand how biogeochemical data correlate with environmental setting. The total range of variation among Roper pyrites is about 70‰,

similar to many Phanerozoic basins^{7,15,16}. In the Roper basin, however, no one facies displays the full range of variation observed for the succession as a whole. Rather, as shown in Fig. 2, different sedimentary facies along the onshore–offshore gradient exhibit consistent differences in $\delta^{34}\text{S}$, with the most negative values occurring in basal shales and the most positive in inner-shelf samples.

Figure 3 shows more clearly the nature and origin of these facies differences. Basinal samples exhibit a wide spread of $\delta^{34}\text{S}$ values, from +20 to +25.5‰ that approximate the isotopic composition of sulphate in Roper seawater¹⁷ to values as negative as –20.7‰ that match the maximum fractionation directly associated with dissimilatory sulphate reduction ($\sim 45\%$)^{3,16}. Note, however, that strong fractionation is evident only in samples whose Fe speciation data indicate euxinic deposition.

In the presence of non-limiting sulphate (perhaps 1 mM, but possibly less), pure cultures and natural populations of sulphate-reducing bacteria fractionate sulphur isotopes strongly, up to 45‰ (refs 3, 16). The marked ^{34}S depletion in basal samples are thus consistent with pyrite deposition from seawater sulphate with an S isotopic value of +20‰ to +25‰ (ref. 17). In this interpretation, both anoxia and the ready availability of sulphate characterize basal waters most likely to have been in direct contact with the global ocean. Within underlying sediments, however, low diffusion rates and high bacterial sulphate demand can deplete supplies of sulphate, with a concomitant reduction in fractionation¹⁸, explaining the few strongly positive $\delta^{34}\text{S}$ values for basal sediments in our study (Fig. 2) and in highly organic-rich shales of the Velkerri Formation¹⁹.

Shallower basinal, distal shelf, and most proximal shelf samples display a consistent range of variation between about +10.2‰ and +27.6‰ (Fig. 2). Strongly positive $\delta^{34}\text{S}$ values, up to +49.7‰, are restricted to shales in coastal environment. The same pattern of facies variation is seen whether one compares similar stratigraphic intervals in different parts of the basin (Urapunga 4 and Golden Grove 1) or different stratigraphic intervals (Urapunga 4 and 5) (Fig. 3). Given the iron speciation chemistry of these samples, shelf pyrites must have formed within organic-rich sediments during early diagenesis. Within sediments, sulphate abundance can become limited by diffusion and as a result, $\delta^{34}\text{S}$ values of early diagenetic pyrites commonly span the full range with strong to essentially no fractionation relative to seawater sulphate^{12,15}. For example, in the modern Black Sea, with a sulphate concentration of about 20 mM, maximum S isotopic fractionations between sedimentary pyrites and seawater sulphate from estuarine, oxic shelf, and euxinic deep-

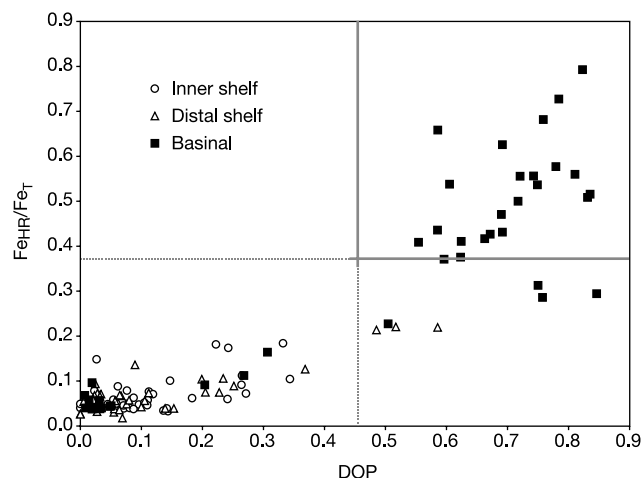


Figure 1 The relationship between the $\text{Fe}_{\text{HR}}/\text{Fe}_{\text{T}}$ ratio and DOP in black shales of the Roper Group. The vertical and horizontal lines cross at a DOP value of 0.45 and a $\text{Fe}_{\text{HR}}/\text{Fe}_{\text{T}}$ ratio of 0.38. In modern marine environments, samples with DOP and $\text{Fe}_{\text{HR}}/\text{Fe}_{\text{T}}$ above these values (within the quadrant enclosed by solid lines) deposit beneath euxinic waters.

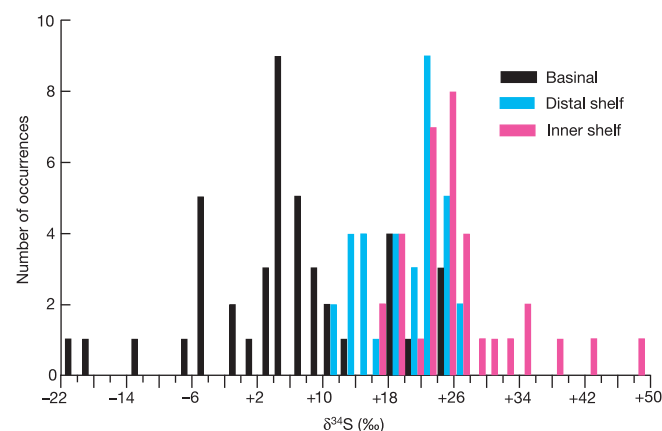


Figure 2 The S isotopic composition of sedimentary pyrites from Roper black shales. Note that samples from different environments form coherent suites. Black, basinal facies; green, distal shelf; pink, inner shelf. $\delta^{34}\text{S}$ in ‰ = $[(^{34}\text{S}/^{32}\text{S})_{\text{sample}} / (^{34}\text{S}/^{32}\text{S})_{\text{standard}} - 1] \times 1000$ relative to the Cañon Diablo Troilite standard.

sea sediments are comparable in magnitude, but oxic environments record a much greater range of values^{13,20}.

The same is true for Phanerozoic examples. Gauthier²¹ measured $\delta^{34}\text{S}$ values for pyrites deposited beneath oxic and anoxic bottom waters of the Cretaceous seas that flooded North America. Like our Roper samples, Cretaceous pyrites from anoxic environments are highly ^{34}S -depleted with strong fractionations relative to seawater sulphate. Unlike the Roper samples, however, most Cretaceous

pyrites that formed within sediments beneath oxic waters are also significantly depleted in ^{34}S with marked fractionations relative to sulphate. The probability that the Roper shelf pyrites were drawn from a sample with a distribution of $\delta^{34}\text{S}$ values like the Cretaceous seaway is vanishingly small ($P < 0.001$; Kolmogorov–Smirnov test).

The observed palaeoenvironmental gradient of S isotopic compositions (Fig. 2) is consistent with previous hypotheses of a sulphate-minimum zone in a stratified Proterozoic ocean²²; such zonation could occur only if marine sulphate levels were low²². Under these conditions, we would expect ^{34}S -enrichment of sulphate within the sulphate-minimum zone, leading to the formation of ^{34}S -enriched sulphides, as observed in Roper shelf samples (Fig. 2). Moreover, in an inner-shelf area, high rates of primary production and rapid sedimentation would further have exaggerated the consequences of low sulphate levels. Particularly under conditions of low sulphate abundance, sulphate removal by bacteria would have outstripped rates of sulphate supply, producing pore-water sulphate (and, in consequence, sulphides) with $\delta^{34}\text{S}$ values much heavier than mean seawater sulphate^{15,16} (Fig. 2). It is hard to produce explanations for the observed facies dependence of S isotopic compositions (Fig. 2) that do not require sea water with sulphate concentrations far below those seen today.

Individual basins need not mirror the chemistry of the oceans as a whole. Therefore, data from other basins are needed to choose unequivocally between local and global interpretations of basinal anoxia and sulphate concentration. Fe speciation data and strongly ^{34}S -enriched pyrites from black shales deposited during maximum flooding of the Tawallah and McArthur basins (about 1,730 Myr and 1,637 Myr ago, respectively), both antecedent to Roper deposition in northern Australia, show deep-water anoxia as well as moderate-to-pronounced sulphate depletion¹². Thus, in this one area of the world, low sulphate and basinal euxinic conditions were present in three successive basins spanning more than 250 million years of Proterozoic Earth history.

The high $\delta^{34}\text{S}$ values of disseminated pyrite in broadly contemporaneous (~1,470–1,440 Myr ago) shales of the Newland Formation, Lower Belt Supergroup, Montana, also indicate sulphate limitation²³. The Belt Basin may have been partially isolated from the global ocean, but more highly fractionated S in pyrites ($\delta^{34}\text{S} = -14\text{‰}$ to $+18\text{‰}$) from the upper Newland Formation suggests at least episodic connection of Belt waters to the Proterozoic ocean²⁴. $\delta^{34}\text{S}$ values of bedded CaSO_4 evaporites from a ~1,200-Myr-old carbonate-evaporite succession in northern Canada vary markedly with stratigraphy; this variation could be explained as follows. Where seawater sulphate levels were low, $\delta^{34}\text{S}$ values for sulphate varied as a function of sulphate removal by bacterial sulphate reduction and consequent iron sulphide precipitation²⁵. In contrast to the Phanerozoic record, positive $\delta^{34}\text{S}$ values characterize early diagenetic pyrites from mid-Proterozoic shelf shales around the world^{15,16}, consistent with low sulphate throughout the world's oceans. Limited data suggest that marine sulphate levels may have remained low until near the end of the Proterozoic eon²⁶.

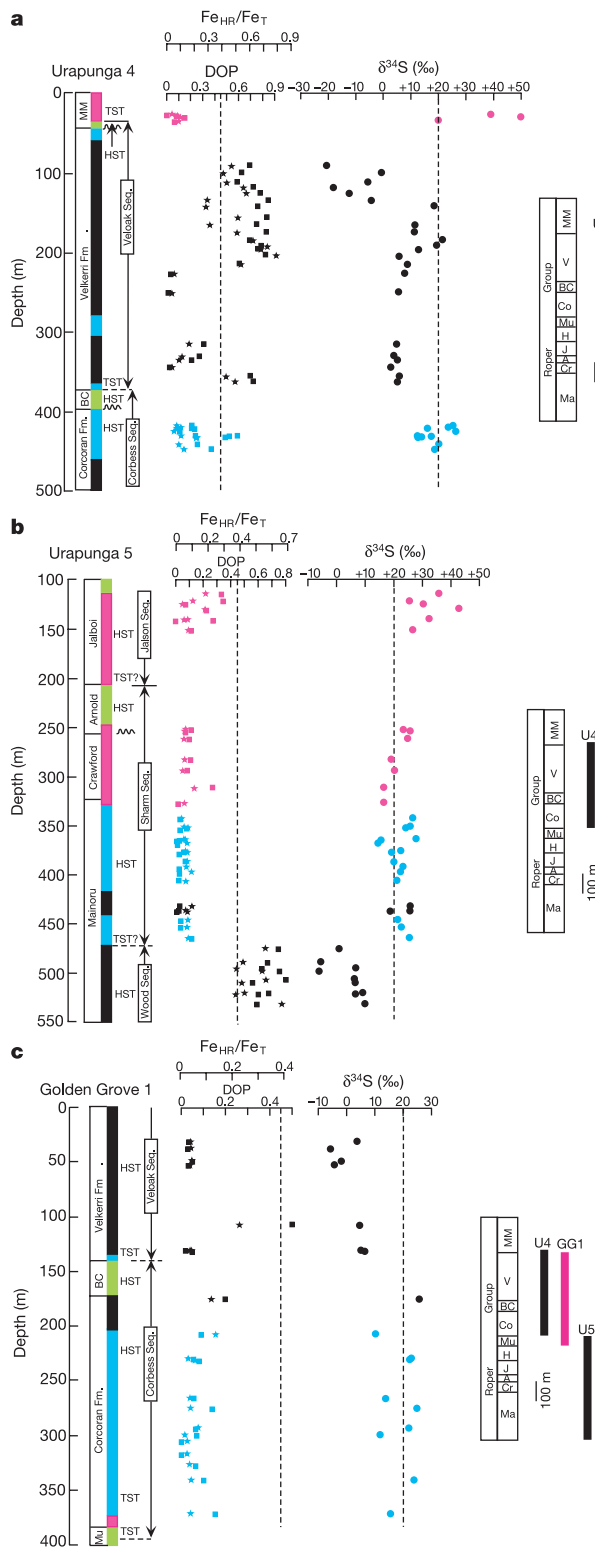


Figure 3 Fe-speciation, S-isotopic, and sequence-stratigraphic data for the Roper Group. Panels **a–c** show data obtained from three drilling sites: **a**, Urupunga 4; **b**, Urupunga 5; **c**, Golden Grove 1. $+20\text{‰}$ represents the possible isotopic composition of open seawater sulphate for middle Proterozoic. In each panel, the left-hand dashed line is drawn at a DOP value of 0.45 (with DOP data shown as filled squares) and an $\text{Fe}_{\text{HR}}/\text{Fe}_{\text{T}}$ ratio of 0.38 (stars). Sequence stratigraphy is from ref. 9. On the right side of each panel, stratigraphy of the Roper Group shows intervals intersected by the drill cores of Urupunga 4 (U4), POG Golden Grove 1 (GG1) and BMR Urupunga 5 (U5). MM, McMinn Formation; V, Velkerri Formation; BC, Bessie Creek Sandstone; Co, Corcoran Formation; Mu, Munyi Member; H, Hodgson Sandstone; J, Jalboi Formation; A, Arnold Sandstone; Cr, Crawford Formation; Ma, Mainoru Formation. Black, basinal facies; green, distal shelf; pink, inner shelf; olive, estuarine. TST, Transgressive Systems Tract; HST, Highland Systems Tract.

The documentation of euxinic and low sulphate conditions in mid-Proterozoic marine basins paves the way to an improved understanding of early life and environments. In such oceans, methanogenic archaeans could have played an enhanced role in the carbon cycle, contributing to long-lived greenhouse conditions²⁷. Low sulphate may also help to explain the prominence of penecontemporaneous dolomite in mid-Proterozoic and older carbonate platforms²⁸. Through its effects on biologically important trace elements, seawater chemistry may help to explain the ecological and evolutionary distributions of early eukaryotic photoautotrophs⁸. And, if sulphate levels remained low until the latest Proterozoic, oxygen probably also remained well below present levels, influencing the early diversification of animals²⁹. Further biogeochemical research carried out in the framework of sequence stratigraphy³⁰ should clarify the relationships between seawater chemistry and changes in the Proterozoic biosphere. □

Received 14 November 2002; accepted 10 April 2003; doi:10.1038/nature01651.

- Holland, H. D. in *Early Life on Earth* (ed. Bengtson, S.) 237–244 (Columbia Univ. Press, New York, 1994).
- Farquhar, J., Bao, H. & Thiemens, M. H. Atmospheric influence of the Earth's earliest sulphur cycle. *Science* **289**, 756–758 (2000).
- Canfield, D. E., Habicht, K. S. & Thamdrup, B. The Archean sulfur cycle and the early history of atmospheric oxygen. *Science* **288**, 658–661 (2000).
- Des Marais, D. J., Strauss, H., Summons, R. E. & Hayes, J. M. Carbon isotopic evidence for the stepwise oxidation of the Precambrian environment. *Nature* **359**, 605–609 (1992).
- Derry, L. A., Kaufman, A. J. & Jacobsen, S. B. Sedimentary cycling and environmental change in the late Proterozoic: evidence from stable and radiogenic isotopes. *Geochim. Cosmochim. Acta* **56**, 1317–1329 (1992).
- Canfield, D. E. & Teske, A. Late Proterozoic rise in atmospheric oxygen concentrations inferred from phylogenetic and stable-isotope studies. *Nature* **382**, 127–132 (1996).
- Canfield, D. E. A new model for Proterozoic ocean chemistry. *Nature* **396**, 450–453 (1998).
- Anbar, A. D. & Knoll, A. H. Proterozoic ocean chemistry and evolution: A bioinorganic bridge? *Science* **297**, 1137–1142 (2002).
- Abbott, S. T. & Sweet, I. P. Tectonic control on third-order sequences in a siliciclastic ramp-style basin: an example from the Roper Superbasin (Mesoproterozoic), northern Australia. *Aust. J. Earth Sci.* **47**, 637–657 (2000).
- Raiswell, R. & Canfield, D. E. Sources of iron for pyrite formation in marine sediments. *Am. J. Sci.* **298**, 219–245 (1998).
- Raiswell, R., Newton, R. & Wignall, P. B. An indicator of water-column anoxia: resolution of biofacies variations in the Kimmeridge clay (upper Jurassic, U.K.). *J. Sedim. Res.* **71**, 286–294 (2001).
- Shen, Y., Canfield, D. E. & Knoll, A. H. Middle Proterozoic ocean chemistry: Evidence from the McArthur Basin, northern Australia. *Am. J. Sci.* **302**, 81–109 (2002).
- Wijsman, J. W. M., Middelburg, J. J., Herman, P. M. J., Böttcher, M. E. & Heip, C. H. R. Sulfur and iron speciation in surface sediments along the northwestern margin of the Black Sea. *Mar. Chem.* **74**, 261–278 (2001).
- Jackson, M. J. & Raiswell, R. Sedimentology and carbon-sulphur geochemistry of the Velkerri Formation, a mid-Proterozoic potential oil source in northern Australia. *Precamb. Res.* **54**, 81–108 (1991).
- Hayes, J. M., Lambert, I. B. & Strauss, H. in *The Proterozoic Biosphere: A Multidisciplinary Study* (eds Schopf, J. W. & Klein, C.) 129–134 (Cambridge Univ. Press, Cambridge, 1992).
- Canfield, D. E. & Raiswell, R. The evolution of the sulfur cycle. *Am. J. Sci.* **299**, 697–723 (1999).
- Muir, M. D., Donnelly, T. H., Wilkins, R. W. T. & Armstrong, K. J. Stable isotope, petrological, and fluid inclusion studies of minor mineral deposits from the McArthur Basin: implications for the genesis of some sediment-hosted base metal mineralization from the Northern Territory. *Aust. J. Earth Sci.* **32**, 239–260 (1985).
- Zaback, D. A., Pratt, L. M. & Hayes, J. M. Transport and reduction of sulfate and immobilization of sulfide in marine black shales. *Geology* **21**, 141–144 (1993).
- Donnelly, T. H. & Crick, I. H. Depositional environment of the middle Proterozoic Velkerri Formation in northern Australia: geochemical evidence. *Precamb. Res.* **42**, 165–172 (1988).
- Lyons, T. W. Sulfur isotopic trends and pathways of iron sulphide formation in upper Holocene sediments of the anoxic Black Sea. *Geochim. Cosmochim. Acta* **61**, 3367–3382 (1997).
- Gautier, D. Cretaceous shales from the western interior of North America: sulfur/carbon ratios and sulfur isotope composition. *Geology* **14**, 225–228 (1986).
- Logan, G. A., Hayes, J. M., Hieshima, G. B. & Summons, R. E. Terminal Proterozoic reorganization of biogeochemical cycles. *Nature* **376**, 53–56 (1995).
- Lyons, T. W., Luepke, J. J., Schreiber, M. E. & Zeig, G. A. Sulfur geochemical constraints on Mesoproterozoic restricted marine deposition: lower Belt Supergroup, northwestern United States. *Geochim. Cosmochim. Acta* **64**, 427–437 (2000).
- Strauss, H. & Schieber, J. A sulfur isotope study of pyrite genesis: The Mid-Proterozoic Newland Formation, Belt Supergroup, Montana. *Geochim. Cosmochim. Acta* **54**, 197–204 (1990).
- Kah, L. C., Lyons, T. W. & Chesley, J. T. Geochemistry of a 1.2 Ga carbonate-evaporite succession, northern Baffin and Bylot Islands: implications for Mesoproterozoic marine evolution. *Precamb. Res.* **111**, 203–234 (2001).
- Hurtgen, M. T., Arthur, M. A., Suits, N. S. & Kaufman, A. J. The sulfur isotopic composition of Neoproterozoic seawater sulfate: implications for a snowball Earth? *Earth Planet. Sci. Lett.* **203**, 413–429 (2002).
- Pavlov, A. A., Hurtgen, M. T., Kasting, J. F. & Arthur, M. A. Methane-rich Proterozoic atmosphere? *Geology* **31**, 87–90 (2003).

- Baker, P. A. & Kastner, M. Constraints on the formation of sedimentary dolomite. *Science* **213**, 214–216 (1981).
- Knoll, A. H. & Carroll, S. B. Early animal evolution: emerging views from comparative biology and geology. *Science* **284**, 2129–2137 (1999).
- Ross, G. M., Bloch, J. D. & Krouse, H. R. Neoproterozoic strata of the southern Canadian Cordillera and the isotopic evolution of seawater sulfate. *Precamb. Res.* **73**, 71–99 (1995).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Bauer and J. Brooks for help in sample collection, J. Fong and L. Liu for help with geochemical analysis, J. Payne for help with statistics, and D. Canfield, J. Hayes, M. Hurtgen, T. Lyons and G. Ross for comments and suggestions. This study was supported by an NRC Research Associateship (Y.S.), the NASA Exobiology programme, the Astrobiology Institute, Macquarie University, and the Australian Research Council.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y.S. (yshen@oeb.harvard.edu).

Stability of forest biodiversity

James S. Clark & Jason S. McLachlan

Center on Global Change, Biology, and Nicholas School of the Environment,
Duke University, Durham, North Carolina 27708, USA

Two hypotheses to explain potentially high forest biodiversity have different implications for the number and kinds of species that can coexist and the potential loss of biodiversity in the absence of speciation. The first hypothesis involves stabilizing mechanisms, which include tradeoffs between species in terms of their capacities to disperse to sites where competition is weak^{1–4}, to exploit abundant resources effectively^{5,6} and to compete for scarce resources⁷. Stabilization results because competitors thrive at different times and places. An alternative, ‘neutral model’ suggests that stabilizing mechanisms may be superfluous. This explanation emphasizes ‘equalizing’ mechanisms⁸, because competitive exclusion of similar species is slow. Lack of ecologically relevant differences means that abundances experience random ‘neutral drift’, with slow extinction^{9–11}. The relative importance of these two mechanisms is unknown, because assumptions and predictions involve broad temporal and spatial scales. Here we demonstrate that predictions of neutral drift are testable using palaeodata. The results demonstrate strong stabilizing forces. By contrast with the neutral prediction of increasing variance among sites over time, we show that variances in post-Glacial tree abundances among sites stabilize rapidly, and abundances remain coherent over broad geographical scales.

It is difficult to test competing hypotheses for coexistence. In ecological models, stable coexistence typically requires tradeoffs. Communities may be ‘packed’, in the sense that only species with particular life history combinations can coexist³. The potential importance of stabilizing forces has led to searches for tradeoffs that partition advantages in space and time between different species¹². Although some types of tradeoffs are clear-cut, such as those characterizing early and late successional species^{13,14}, the overall importance of tradeoffs is debatable^{10,12,15}.

By contrast, if species differences are small, the inherent limits to diversity implied by specific types of competition may not apply. The neutral model does not refer to a strict random walk, because finite space limits the density of trees. Rather, ecological equivalence allows that abundances drift randomly. Speciation might balance losses, provided that new species are generated as rapidly as random extinction removes them. Although there are unique temporal patterns associated with processes such as random drift, for trees

the dynamics are too slow to quantify adequately. Fossil pollen records represent time series of species abundance, but statistical tests for a random walk (see for example ref. 16) are not powerful. Thus, the case for lack of tradeoffs typically appeals to indirect evidence, such as the obvious similarities shared by many tree species¹⁰, and spatial patterns or dominance–diversity curves¹¹ that might be compatible with multiple mechanisms. Hubbell¹¹ suggests that, even if wrong, difficult to test, or both, the hypothesis of random drift might serve as a useful ‘null’ model.

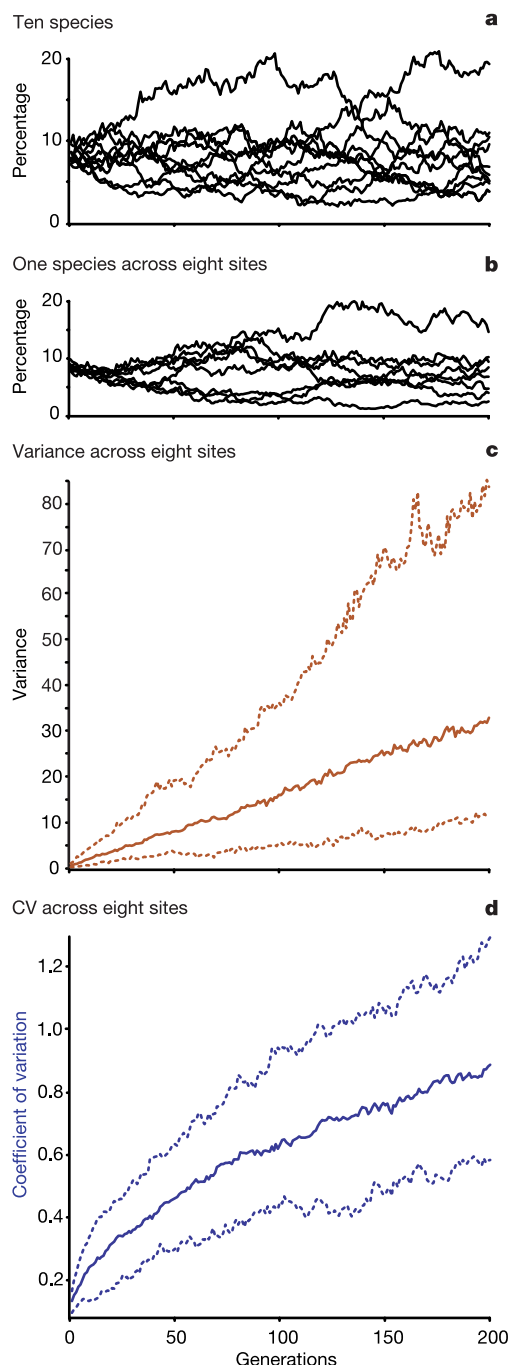


Figure 1 Simulations of neutral dynamics showing the divergence among sites. **a**, Lottery model, where ten species with identical parameters and responses to stochasticity compete for space. **b**, Results for a single species shown for eight different sites. Abundances diverge with the random accumulation of changes at each site. **c**, Variance among sites increases over time owing to accumulation of the random changes in abundance shown in **b**, as does **(d)** the coefficient of variation, CV. In **c** and **d**, the middle line indicates the median, and the dashed lines bound 90% of simulated values.

Tests of neutral drift are available, provided that we exploit both temporal and spatial information from fossil pollen data, as opposed to trends in an individual series. If community dynamics are characterized by random drift, then variances among sites increase over time, with the accumulation of random changes in abundance. This is true even if the process is not strictly neutral (see Methods). For example, Fig. 1a illustrates a lottery process for species having identical parameters, where responses to stochasticity do not depend on species identity. To stress the fact that variance among sites increases rapidly over time, our simulation example assumes within-site variances that are substantially lower than we calculate from fossil data. Drift results in rapid accumulation of among-site variance (Fig. 1b, c), which is readily identifiable and would continue to increase until much of the diversity was lost through extinction. This trend is also predicted for the coefficient of variation (Fig. 1d), which, unlike the variance, does not increase simply because absolute abundance increases.

We analysed trends in variance among sites from a portion of eastern North America during the Holocene epoch. Fossil pollen that accumulates in lake sediments records changes in the relative abundances of species over geological timescales, but with temporal resolution that is compatible with tree generation times (10 to ~100 years). The fossil pollen record integrates abundances over scales of a ‘stand’ (small lakes) to a region (large lakes)^{17,18}. Sites used here are distributed over more than 10⁴ km², an area large enough that dynamics are mutually independent; rare, long-distance dispersal is not sufficient to synchronize dynamics between sites. Thus, it represents a useful scale for assessing the degree to which populations may engage in neutral dynamics (see Methods).

To test the hypothesis that abundances drift at random, we determined changes in means and variances for the past 10,000 yr (as dated by ¹⁴C), a period of more than 200 tree generations, at eight sites. This interval spans a time of rapid climate change followed by modest climate changes typical of those prevailing in forests postulated to show neutral drift. The interval is not free of external influences. Indeed the rapid climate changes at the beginning of the Holocene and the large-scale changes in hemlock (*Tsuga*) abundance (probably caused by a natural enemy¹⁹) lend additional insight into the neutral hypothesis. Climate fluctuations contribute variance to populations. Failure to observe the rise in variance predicted by neutral dynamics would be especially strong evidence for stabilizing forces, which would have to overcome the tendency of climate variability, disturbance and the hemlock decline to increase variability. We used regression to assess trends in among-site variance over time.

Our results are not consistent with neutral drift. Contrary to predictions (Fig. 1), following expansion in the early Holocene the relative abundances of all dominant pollen taxa reach relatively constant values, and variances and coefficients of variation among sites tend to stabilize or decline (Fig. 2). The expansion persists well into the Holocene at these mid-latitude sites and is characterized by increasing pollen percentages, attended by a rise in among-site variance. The initial increases in among-site variance result from increase in total abundance, not from drift, as is shown by coefficients of variation. The percentages for a given taxon subsequently tend to stabilize at each site. In contrast with neutral drift (Fig. 1c), none of the changes in variance are significant (Fig. 2). Only *Betula* and *Quercus* show evidence of increasing variance during the past few millennia. For both taxa, this trend follows declining variances of the mid-Holocene, and it results from increasing abundances and not from divergence among sites. Moreover, *Acer*, *Fagus*, *Fraxinus* and *Ulmus* all show declining variances among sites in the latter half of the Holocene.

Tsuga canadensis provides especially compelling evidence for stabilizing mechanisms, and in a different way. Following disturbance, stabilizing forces are expected to return populations to their former abundances. By contrast, the neutral model predicts that

population densities should continue to drift at random, with no propensity to return to previous abundances. In all eight cases, the synchronous collapse of this population across eastern North America at 4750 ^{14}C yr BP (ref. 20) was followed by return to abundances close to those prevalent before its decline. The fact that this occurs across eastern North America provides strong evidence that stabilizing mechanisms maintain species abundances, and it is not consistent with neutral drift.

We selected a specific region for our analysis for purposes of demonstration, but the results can be generalized. Pollen data from temperate and boreal regions that have abundant pollen records share the characteristics of geographical coherence and continuity^{21,22}. We could not complete this analysis for tropical forests as lake sediments are not readily available. Nonetheless, relatively constant pollen abundances spanning much of the Holocene are typical of the Tropics²³, consistent with our results and indicating that stabilizing mechanisms are not limited to temperate latitudes.

Our results provide strong evidence for stabilizing mechanisms. Populations did not achieve random densities following deglaciation at the end of the Pleistocene and neutrally drift thereafter. Rather, the strong geographical coherence in early Holocene pollen abundances and broad consistency in variances among sites point to stabilizing effects. Even the collapse of a dominant late-successional species, hemlock, saw recovery to relative abundances typical of pre-collapse populations.

Taken together, these aspects of the palaeo-record point to the

importance of stabilizing mechanisms that maintain tree populations within bounded densities which are geographically and temporally coherent. This evidence does not, in itself, demonstrate a role for tradeoffs in any specific sense. It does indicate the importance of stabilizing mechanisms that might derive from species differences, in terms of their capacities to colonize new sites, exploit abundant resources and tolerate competition^{3–8}. Individual-level variability can result in large overlap among species, but the effect can be stabilizing^{12,15}. Stochasticity associated with colonization, growth and survival can have a strong stabilizing influence that contrasts with the slow drift to extinction or dominance implied by neutral dynamics^{8,12,15}. Our results do not imply that the equalizing mechanisms stressed in the neutral model are absent or inconsequential^{9–11}, only that they must operate together with strong stabilizing forces.

A neutral interpretation of forest dynamics requires that continual (albeit potentially slow) extinction losses are offset by persistent speciation^{9,11}. This implies that modern diversity levels owe much to continual emergence of new species. We did not test the neutral drift model using speciation and extinction data for two reasons. First, speciation and extinction rates are invoked to explain why diversity does not eventually collapse with neutral drift, but neutral drift can be tested on its own. Speciation and extinction can occur whether dynamics are controlled by stabilizing forces or by neutral drift. Second, the palaeo-record is inadequate for tests of speciation and extinction. Whereas patterns of extinction and

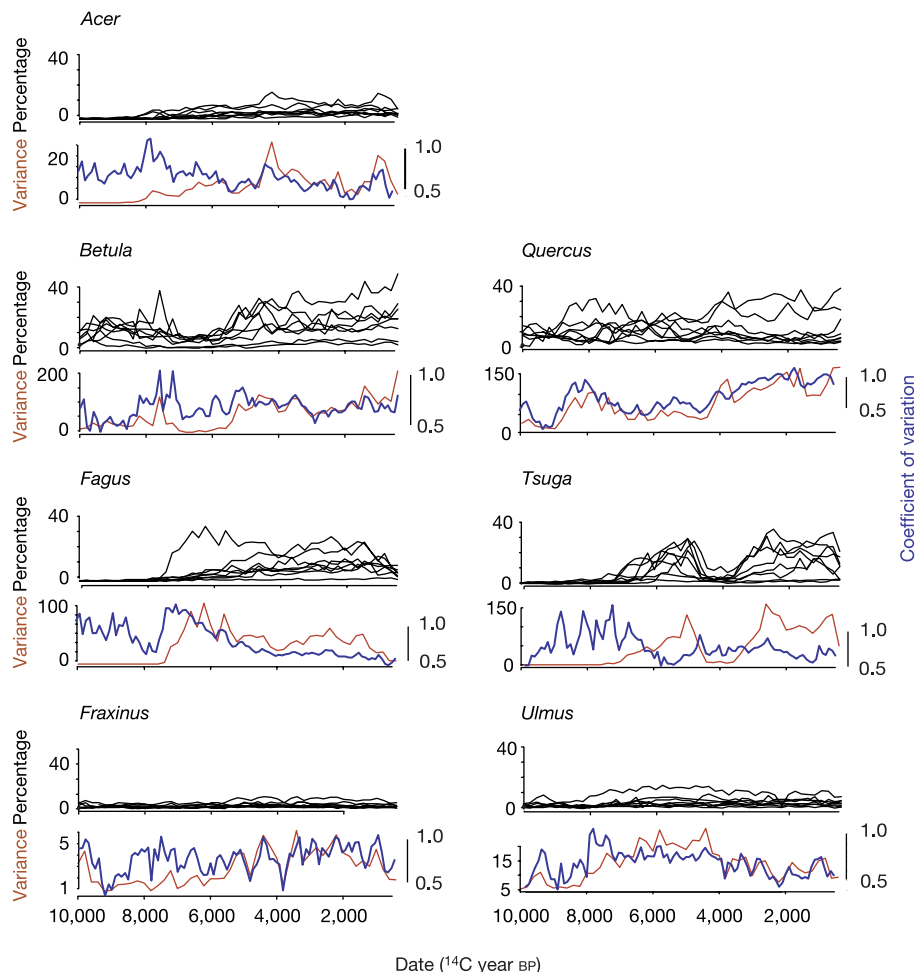


Figure 2 Fossil pollen data for seven dominant taxa from eight sites in Ontario. In contrast to 'neutral' changes in **a**, abundances of all taxa at all sites tend to stabilize soon after expansion. Although the variances within each series are much larger than those used for

simulation, the variances (brown) and coefficients of variation (blue) among sites do not increase.

speciation are poorly known and, by themselves, do not identify mechanisms, geographical patterns in variance are compelling. Our analysis demonstrates that drift is inconsistent with evidence that is readily available for the Holocene.

The strong evidence for stabilizing forces in the palaeo-record has implications different from those of neutral drift. If modern diversity levels have been maintained by stabilizing forces, rather than by incessant replenishment through speciation, then high contemporary extinction rates are likely to have more profound and enduring impacts on future biodiversity. □

Methods

Data

Fossil pollen percentages were obtained from the North American Pollen Database (<http://www.ngdc.noaa.gov/paleo/pollen.html>). The eight sites come from southern Ontario, where there exists an unusually high density of small lakes (Table 1). These sites are small enough to emphasize local dynamics, they share regional climate and vegetation histories, and they are spaced broadly enough to allow comparisons among sites. To calculate variances among sites, we generated evenly spaced series at 200-yr intervals by linear interpolation from 10,000 to 200 ¹⁴C yr BP. Two of the lakes show evidence of pre-Columbian agriculture (Second and Tonawa Lakes), which could affect tree abundances, but omission of these sites did not change the results.

Pollen data are not the same as tree data, and the relationships between them influenced our modelling approach. Pollen abundances are sometimes fitted to models of population dynamics. Pollen accumulation rate is typically used with such applications, and it is compatible with models of population growth rate. Those models can be interpreted in terms of either population density or abundance. The lottery model is appropriate for assessment of neutral dynamics, and it assumes relative abundance. We therefore use pollen percentages, not pollen accumulation. Because different taxa produce different amounts of pollen, the relative abundance of pollen is not the same as the relative abundance of trees. Here again, a lottery model is appropriate, because the predicted trends in variance hold even for arbitrary weightings of species. The fact that some pollen taxa include more than one species does not affect the assessment of this mechanism. We could view pollen percentages as the sums of random variables, and the spatial variance among those sums would also increase with time. Declines in variance, coefficient of variation, or both for the monospecific taxa (*Fagus*, *Tsuga*) confirm that individual species also show strong evidence of stabilizing forces.

Simulation model

‘Neutrality’ can be viewed in several ways. Regardless of the specific model, we simply point out that abundances among sites tend to diverge over time. This is true whether the process is a simple random walk or there is some constraint on total density. A random walk on log density could be written as

n_{ij}(t+1) = n_{ij}(t)e^{e_{ij}(t)}

for density n_{ij}(t) of species i at location j at time t, and stochasticity might be represented as e_{ij}(t) ~ N(0, sigma^2), with variance sigma^2. Variance among sites j increases linearly with time. In this model, there are no bounds on total density. Density-dependent versions of this model²⁴ also show divergence across sites.

Some descriptions of ‘neutrality’ are better represented by a lottery model than by a random walk, because the space available is finite. To illustrate trends in abundance among sites j, consider a simple lottery

n_{ij}(t+1) = (1 - delta_i)n_i(t) + n_i(t)sum_{k=1}^m delta_k n_{kj}(t) / sum_{k=1}^m beta_k(t) n_{kj}(t) (1)

with mortality rate delta_i and stochastic recruitment rate beta_i(t) ~ Pois(b_i). The first term in equation (1) is survival. The second term is new recruitment, being the product of new sites opened by mortality and the fraction of those sites that are obtained by species i. Although the state variable n(t) is continuous, our discrete recruitment allows random changes in population to have an effect at low density. Stable coexistence results, provided that species differ in how they respond to stochasticity, which, in this case, is associated with recruitment.

Neutrality implies not only that parameters are identical for all species

delta_i = delta and b_i = b, i = 1, ..., m

but also that they respond identically to stochasticity. We can remove distinctions among

individuals in terms of species identity with the following modification:

n_{ij}(t+1) = (1 - delta)n_i(t) + n_i(t)delta sum_{k=1}^m n_{kj}(t) / sum_{k=1}^m beta_{kj}(t) (2)

for stochastic recruitment rate beta_{ij}(t) ~ Pois(bn_{ij}(t)). To see how this removes species identities, consider a neutral recruitment process, whereby each individual contributes Pois(b). Provided that individuals k are independent, then total propagule production is the random variate distributed as sum_{i=1}^m sum_{k=1}^{n_{ij}} Pois(b) ~ Pois(b sum_{i=1}^m n_{ij}(t)) ~ sum_{i=1}^m Pois(bn_{ij}(t)), and production by species i is distributed as Pois(bn_{ij}(t)). Equation (2) produces slow drift as described by a number of authors (see for example ref. 9).

Obviously, the increase in variance across sites that we mention above applies in this case. Of course, all stochastic models initialized at identical densities will show some initial increase in variance among sites. Our examples use initial variability comparable to differences that we observe among sites in the early Holocene (Fig. 1a). The variance plot (Fig. 1c) bounds 90% of variance trends for 100 random initializations of the neutral lottery model. Coefficients of variation indicate when changes in variance occur because of changing overall abundance.

Received 10 February; accepted 2 April 2003; doi:10.1038/nature01632.

1. Horn, H. S. & MacArthur, R. H. Competition among fugitive species in a harlequin environment. *Ecology* **53**, 749–752 (1972).

2. Hastings, A. Disturbance, coexistence, history, and competition for space. *Theor. Pop. Biol.* **18**, 363–373 (1980).

3. Tilman, D. Competition and biodiversity in spatially structured habitats. *Ecology* **75**, 2–16 (1994).

4. Hixon, M. A., Pacala, S. W. & Sandin, S. A. Population regulation: historical context and contemporary challenges of open vs. closed systems. *Ecology* **83**, 1490–1508 (2002).

5. Pacala, S. W. & Rees, M. Field experiments that test alternative hypotheses explaining successional diversity. *Am. Nat.* **152**, 729–737 (1998).

6. Bolker, B. M. & Pacala, S. W. Spatial moment equations for plant competition: understanding spatial strategies and the advantage of short dispersal. *Am. Nat.* **153**, 575–602 (1999).

7. Tilman, D. *Plant Strategies and the Dynamics and Structure of Plant Communities* (Princeton Univ. Press, Princeton, New Jersey, 1998).

8. Chesson, P. Mechanisms of maintenance of species diversity. *Annu. Rev. Ecol. Syst.* **31**, 343–366 (2000).

9. Hubbell, S. P. & Foster, R. B. in *Community Ecology* (eds Diamond, J. & Case, T. J.) 314–330 (Harper and Row, New York, New York, 1986).

10. Hubbell, S. P. *et al.* Light gap disturbances, recruitment limitation and tree diversity in a neotropical forest. *Science* **283**, 554–557 (1999).

11. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, New Jersey, 2001).

12. Clark, J. S., Dietze, M., Ibanez, I. & Mohan, J. Coexistence: how to discriminate trophic tradeoffs from slow dynamics. *Ecology* (in the press).

13. Huston, M. & Smith, T. Plant succession: life history and competition. *Am. Nat.* **130**, 168–198 (1987).

14. Rees, M., Condit, R., Crawley, M., Pacala, S. & Tilman, D. Long term studies of vegetation dynamics. *Science* **293**, 650–655 (2001).

15. Clark, J. S., LaDeau, S. & Ibanez, I. Fecundity of trees and the colonization–competition hypothesis. *Ecological Monogr.* (in the press).

16. Dickey, D. A. & Fuller, W. A. Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.* **74**, 427–431 (1979).

17. Jacobson, G. L. & Bradshaw, R. H. W. The selection sites for paleovegetational studies. *Quat. Res.* **16**, 80–96 (1981).

18. Prentice, I. C. Pollen representation, source area, and basin size: toward a unified theory of pollen analysis. *Quat. Res.* **23**, 76–86 (1985).

19. Allison, T. D., Moeller, R. E. & Davis, M. B. Pollen in laminated sediments provides evidence for a mid-Holocene forest pathogen outbreak. *Ecology* **67**, 1101–1105 (1986).

20. Bennett, K. D. & Fuller, J. L. Determining the age of the mid-Holocene *Tsuga canadensis* (hemlock) decline in eastern North America. *Holocene* **12**, 421–429 (2002).

21. Huntley, B. & Birks, H. J. B. *An Atlas of Past and Present Pollen Maps for Europe 0–13,000 Years Ago* (Cambridge Univ. Press, New York, 1983).

22. Jackson, S. T., Overpeck, J. T., Webb, T., Keatts, S. E. & Anderson, K. H. Mapped plant-macrofossil and pollen records of late Quaternary vegetation change in eastern North America. *Quat. Sci. Rev.* **16**, 1–70 (1997).

23. Bush, M. B., Miller, M. B., De Oliveira, P. E. & Colinvaux, P. A. Two histories of environmental change and human disturbance in eastern lowland Amazonia. *Holocene* **10**, 543–553 (2000).

24. Dennis, B. & Taper, M. L. Density dependence in time series observations of natural populations: estimation and testing. *Ecol. Monogr.* **64**, 205–224.

25. Bennett, K. D. Holocene history of forest trees in southern Ontario. *Can. J. Bot.* **65**, 1792–1801 (1987).

26. Burden, E. T., McAndrews, J. H. & Norris, G. Palynology of Indian and European forest clearance and farming in lake sediment cores from Awenda Provincial Park, Ontario. *Can. J. Earth Sci.* **23**, 43–54 (1986).

27. Fuller, J. L. Holocene forest dynamics in southern Ontario, Canada: fine-resolution pollen data. *Can. J. Bot.* **75**, 1714–1727 (1997).

28. McAndrews, J. H. & Campbell, I. D. in *Proxy Climate Data and Models of the Six Thousand Years Before Present Time Interval: The Canadian Perspective* (ed. Telka, A.) (Canadian Global Change Program Incidental Rep. Series IR93-3, The Royal Society of Canada, Ottawa, 1993).

29. Szeicz, J. M. & MacDonald, G. M. Postglacial vegetation history of oak savanna in southern Ontario. *Can. J. Bot.* **69**, 1507–1519 (1991).

Acknowledgements Research was supported by the National Science Foundation. We thank K. Bennett and P. Chesson for comments.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.S.C. (jimclark@duke.edu).

Table 1 **Lakes and references for sites analysed in Fig. 2**

Site	Latitude (°N)	Longitude (°W)	Elevation (m)	Reference
Lac Bastien	46.24.00	78.55.00	305	25
Decoy Lake	43.14.00	80.22.00	260	26
Graham Lake	45.11.00	77.21.00	381	27
Hams Lake	43.14.12	80.24.48	301	25
High Lake	44.31.00	76.36.00	192	27
Nutt Lake	45.13.00	79.27.00	305	25
Second Lake	44.51.00	79.58.48	196	28
Tonawa Lake	44.51.00	77.10.30	274	29

Scaling metabolism from organisms to ecosystems

Brian J. Enquist^{*,†}, Evan P. Economo^{*}, Travis E. Huxman^{*}, Andrew P. Allen[‡], Danielle D. Ignace^{*} & James F. Gillooly[‡]

^{*} Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA

[†] Center for Applied Biodiversity Science, 1919 M. Street, NW, Suite 600, Washington DC 20036, USA

[‡] Department of Biology, University of New Mexico, Albuquerque, New Mexico 87131, USA

Understanding energy and material fluxes through ecosystems is central to many questions in global change biology and ecology^{1–11}. Ecosystem respiration is a critical component of the carbon cycle^{1,5–7} and might be important in regulating biosphere response to global climate change^{1–3}. Here we derive a general model of ecosystem respiration based on the kinetics of metabolic reactions^{11–13} and the scaling of resource use by individual organisms^{14,15}. The model predicts that fluxes of CO₂ and energy are invariant of ecosystem biomass, but are strongly influenced by temperature, variation in cellular metabolism and rates of supply of limiting resources (water and/or nutrients). Variation in ecosystem respiration within sites, as calculated from a network of CO₂ flux towers^{5,7}, provides robust support for the model's predictions. However, data indicate that variation in annual flux between sites is not strongly dependent on average site temperature or latitude. This presents an interesting paradox with regard to the expected temperature dependence. Nevertheless, our model provides a basis for quantitatively understanding energy and material flux between the atmosphere and biosphere.

Our ability to predict variation in ecosystem processes is currently limited by our ability to mechanistically link biological processes across both spatial and temporal scales^{4,8–10}. One promising approach is to focus on how biotic and abiotic factors regulate metabolic rates of individuals, which combine to determine ecosystem flux rates. Metabolism is the fundamental process dictating material and energy fluxes through organisms^{10–18}. Recent work^{11–14,16} shows that most variation in the metabolic rates of individuals, B_i , can be quantified on the basis of the combined effects of two variables, body size, M_i , and absolute temperature, T in kelvins (K), using the general model for metabolic scaling:

$$B_i = b_0 e^{-E/kT} M_i^{3/4} \quad (1)$$

where b_0 is a normalization constant independent of body size and temperature. The 3/4 power scaling exponent reflects the constraints on resource supply to individual cells through fractal-like distribution networks¹¹. The Boltzmann factor, $e^{-E/kT}$, describes the temperature-dependence of metabolic rate¹², where E is the average activation energy of metabolism (~ 0.6 eV; see ref. 12) and k is Boltzmann's constant (8.62×10^{-5} eV K⁻¹). Previous work indicates that the normalization constant, b_0 , and the activation energy, E , are approximately constant for plants and microbes^{12,13}, the two groups that comprise most of the biomass in terrestrial ecosystems.

Here we build on the general model (equation (1)) to account for variation in rates of ecosystem respiration at sites across the globe. We assume that, at a given site, organisms grow and fill physical space so that the rate of resource use by all individuals, Q_{tot} , is proportional to R , the approximate rate of limiting resource supply (such as water and/or nutrients) when evaluated per unit area per unit time (see ref 10). The respiration rate of an ecosystem, B_e , is equal to the sum of the individual metabolic rates, B_i , for all

Box 1

A general model for scaling biochemical kinetics from organisms to ecosystems

The total ecosystem metabolic flux per unit area, B_e , is influenced by the number of organisms of a given size, M_i , and their respective metabolic rates, B_i . To account for the allometric dependence of B_e , we conduct the summation of B_i across n discrete body size classes, indexed by j , from the smallest sizes (m_1) to the largest sizes (m_n). Here m_j is the average mass within a given arbitrary bin or size class used to resolve the size distribution. Specifically, the whole-system metabolism is the summation of the average metabolic rate of all organisms within each size class, B_j , and their associated total population density, N_j , so that

$$B_e \approx \sum_{j=1}^n [(B_j)(N_j)]$$

Here the total biomass contained within the j th bin is given by $M_j^{\text{tot}} = m_j N_j$ and the density of individuals per bin is given by $N_j = M_j^{\text{tot}} / m_j$. Thus, from equation (1) we have

$$B_e = \left[\sum_{j=1}^n (e^{-E/kT} b_0 m_j^{3/4})(N_j) \right] = e^{-E/kT} b_0 \left[\sum_{j=1}^n (m_j^{3/4}) \left(\frac{M_j^{\text{tot}}}{m_j} \right) \right] \quad (1.1)$$

Equation (1.1) can then be simplified as

$$Q_{\text{tot}} \propto B_e = e^{-E/kT} b_0 \left[\sum_{j=1}^n (m_j^{-1/4})(M_j^{\text{tot}}) \right] \quad (1.2)$$

where the $-1/4$ power accounts for the allometric scaling of organismal mass-specific metabolism. The proportion (α_j) of the total standing ecosystem biomass, M_e^{tot} , that resides within each mass bin, m_j , is $\alpha_j \equiv m_j N_j / M_e^{\text{tot}}$ so that $M_j^{\text{tot}} = \alpha_j M_e^{\text{tot}}$, where $0 < \alpha_j \leq 1$ and the sum

$$\sum_{j=1}^n \alpha_j = 1$$

For a given M_e^{tot} , the value of α_j is influenced by the width of the bin used to resolve the size distribution ($m_j - m_{j-1}$), the density of individuals per bin (N_j) and the total number of bins (n) relative to the total standing biomass, M_e^{tot} .

Simplifying gives the general form of the ecosystem respiration equation,

$$Q_{\text{tot}} \propto B_e = e^{-E/kT} b_0 \left[M_e^{\text{tot}} \left(\sum_{j=1}^n \alpha_j m_j^{-1/4} \right) \right] \quad (1.3)$$

where the term

$$\left(\sum_{j=1}^n \alpha_j m_j^{-1/4} \right)$$

represents the allometric dependence as reflected in the community size distribution. Empirical data and allometric theory show that $Q_{\text{tot}} \approx B_e \approx R$ is independent of the standing biomass, M_e^{tot} (ref. 10). This invariance, or the energetic equivalence rule (EER)^{10,17}, predicts that within a given environment, if R is constant, the value of B_e is independent of M_e^{tot} . Thus, the value of

$$\left[M_e^{\text{tot}} \left(\sum_{j=1}^n \alpha_j m_j^{-1/4} \right) \right]$$

(henceforth shown as C) is constrained to be proportional to resource availability, so that $C \propto R$. Therefore, according to the EER, if R varies between environments (and T , E and b_0 are constant) our model predicts that B_e and Q_{tot} will vary accordingly as $Q_{\text{tot}} \approx B_e \approx C \propto R$. As a result, variation in R must then influence the nature of the size distribution within a given ecosystem through linked or separate variation in M_e^{tot} , the density of individuals of a given size ($m_j N_j$) and the maximum (m_n) and/or minimum (m_1) sizes of organisms.

We can now rearrange equation (1.3) into the following general form:

$$\ln(Q_{\text{tot}}) \propto \ln(B_e) = \frac{-E}{1,000K} \left(\frac{1,000}{T} \right) + \ln[(b_0)(C)] \quad (1.4)$$

where, again, $C \propto R$. Equation (1.4) is similar in form to an Arrhenius plot for calculating activation energies for biochemical reactions^{19,20,32,33}. Here, however, values of E represent the activation energy for metabolism across plants, animals and microbes found within a given assemblage (see Methods and ref. 12).

organisms in that system:

$$R \propto Q_{\text{tot}} \propto B_e = \sum_i B_i \quad (2)$$

This expression can be simplified by using the derivation in Box 1 to yield the following expression:

$$\ln(R) \propto \ln(Q_{\text{tot}}) \propto \ln(B_e) = \frac{-E}{1,000k} \left(\frac{1,000}{T} \right) + \ln[(b_0)(C)] \quad (3)$$

Equation (3) provides a general function for the scaling of ecosystem respiration and yields three important predictions. First, for a given ecosystem, plotting $\ln(B_e)$ against $(1,000/T)$ should yield a general linear relationship describing the dependence of ecosystem respiration on temperature. Second, the slope of this relationship, across diverse ecosystems, should be $-E/1,000k \approx -7.5 \text{ K}$, reflecting the fundamental importance of the activation energy of metabolism in constraining whole-ecosystem respiration (see Methods). Last, because of the energetic equivalence rule (EER; see Box 1), ecosystem respiration should be independent of the total standing biomass across sites if rates of resource supply are held constant (for example, grasslands and forests can have similar values of B_e). Equation (3) also shows the functional

dependence of the intercept on b_0 and C , which characterize the intensity of cellular metabolism, and the size abundance distribution of organisms, respectively (Box 1). Note that the limiting resource supply rate, R , might influence ecosystem respiration through its effects on one or both of these variables.

We evaluated these predictions by compiling nightly ecosystem CO_2 flux rates, B_e , throughout the year from flux towers across the globe. Data were assembled from FLUXNET^{5,7} (see Methods and Supplementary Table 1). Plotting $\ln(B_e)$ (in W ha^{-1}) against the reciprocal of absolute temperature ($1,000/T$) for 19 eddy covariance flux-tower sites in both Europe and North America provides robust support for the model's predictions (Fig. 1; Supplementary Table 1). First, at all sites, the natural logarithm of ecosystem respiration rate is linearly related to the reciprocal of absolute temperature. Second, the average of the observed slopes across ecosystems for all years ($\bar{x} = -7.17$, 95% confidence interval -6.53 to -7.81 , $n = 45$) is indistinguishable from the value predicted from the temperature dependence of individual metabolism (-7.79 ± 1.63) (ref. 12). Furthermore, the slopes all fall within the range expected from the measured activation energies of metabolic reactions (see Methods). Third, analysis of the 95% confidence intervals indicates a strong overlap in the values of the slopes and annual night-time CO_2 fluxes

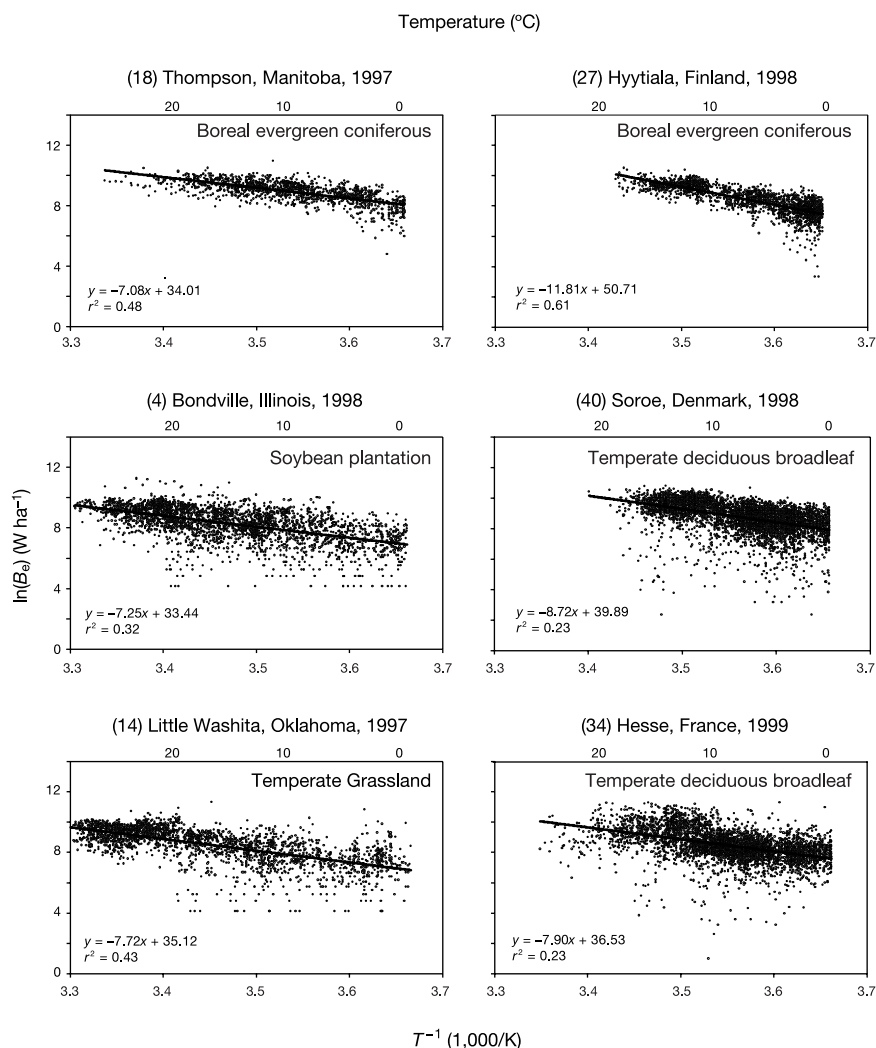


Figure 1 A plot of night-time ecosystem flux, $\ln(B_e)$, against the reciprocal of absolute temperature for 30-min average samples throughout the year for six representative eddy covariance flux-tower sites. The slope of the temperature response reflects the activation

energy of this rate process. Both the range and the average of the slopes across ecosystems are statistically indistinguishable from predictions of the model. Numbers on the top portion of each graph refer to temperature in $^{\circ}\text{C}$.

across all ecosystems, which differ in standing biomass and floristic composition, and which include 14 forests, 2 grasslands, 2 desert sites and an agricultural site (Figs 2a and 3; Supplementary Table 1). The similarity of the slopes of these relationships across sites comprising different taxa (for example, Europe versus America), plant functional types (conifers versus angiosperms), photosynthetic pathways (C_3 versus C_4) and diversity (low latitude versus high latitude) provides strong evidence that ecosystem flux is constrained by the activation energy of individual metabolism.

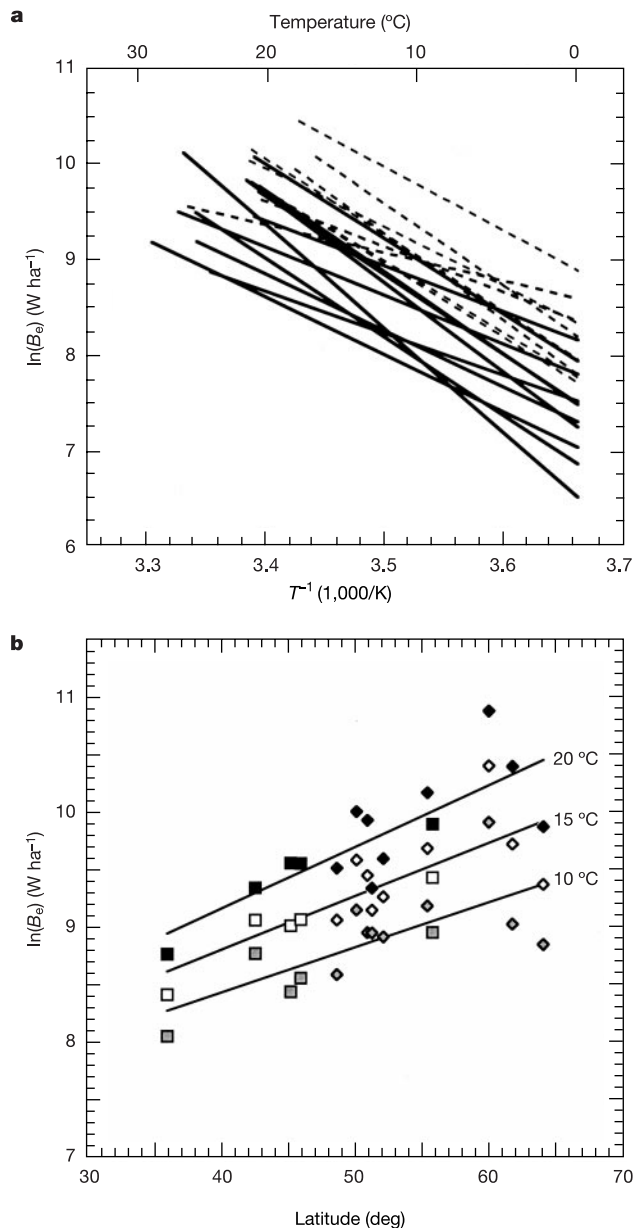


Figure 2 Summary of variability in the temperature and flux function across sites. These figures indicate an increase in the temperature-normalized instantaneous flux with increasing latitude. **a**, Fitted night-time flux functions for all 19 sites in North America (solid lines) and Europe (dashed lines). **b**, Plot of the average ecosystem flux intercept for each forest site normalized at three temperatures (20, 15 and 10 °C; that is, $\ln(B_e(20\text{ °C}))$) against latitude shows a significant positive relationship (diamonds, Europe; squares, North America). All forest sites at 20 °C: $r^2 = 0.651$, $n = 14$, $P = 0.000487$; European sites at 20 °C: $r^2 = 0.361$, $n = 9$, $P = 0.087$; North American sites at 20 °C: $r^2 = 0.915$, $n = 5$, $P = 0.0108$. Black, white and grey symbols represent flux values normalized at 20, 15 and 10 °C respectively.

The empirical fits to the predicted flux functions reveal a wide range of intercept values for European and North American sites (see Fig. 2a). This variation might be due to a systematic change in the temperature-corrected ecosystem respiration rate with latitude². Applying analysis of covariance (ANCOVA) to the entire flux data set, we find a significant overall effect of temperature on ecosystem flux (type III SS: d.f. = 1, $F = 20,888$, $P < 0.0001$). In addition, we find significant differences between sites and years in the intercepts (d.f. = 44, $F = 743.9$, $P < 0.0001$) and slopes (type III SS: d.f. = 44, $F = 63.9$, $P < 0.0001$). Holding the slope constant for all sites (pooled ANCOVA slope of -7.40) still reveals significant differences in the fitted intercepts between sites (type III SS: d.f. = 44, $F = 725.52$, $P < 0.0001$). Such differences might also reflect differences in resource supply (R) between sites. Restricting the analysis to forest sites only, where values of M_e^{tot} are similar (see also ref. 10), still reveals differences in fitted intercepts across sites (ANCOVA constant slopes model type III SS: d.f. = 36, $F = 522.72$, $P < 0.0001$).

The within-site average fitted intercepts (from the fixed-slopes model from the ANCOVA) for all sites and for forest sites are positively correlated with latitude (all sites: $n = 19$, $r^2 = 0.61$, $F = 27.07$, $P = 0.0001$; forest sites: $n = 14$, $r^2 = 0.424$, $F = 8.83$, $P = 0.012$). Thus, geographic variation in the fitted intercepts indicates that high-latitude ecosystems are characterized by an approximately 3–6-fold increase in CO_2 flux at a given temperature when compared with lower-latitude sites. For each forest site, plotting the temperature-normalized flux rates at three different temperatures (20, 15 and 10 °C) shows a significant increase with latitude (Fig. 2b). Perhaps as a consequence, the annual night-time CO_2 flux does not vary significantly with average annual temperature or latitude (Fig. 3; $P > 0.10$; see also ref. 1). There are several possible explanations for this pattern. These include latitudinal gradients in resource availability (R) or community structure (related to C), lag effects from daytime respiration, and calibration of towers across sites. These results are also qualitatively consistent with results indicating that individuals in colder climates adjust

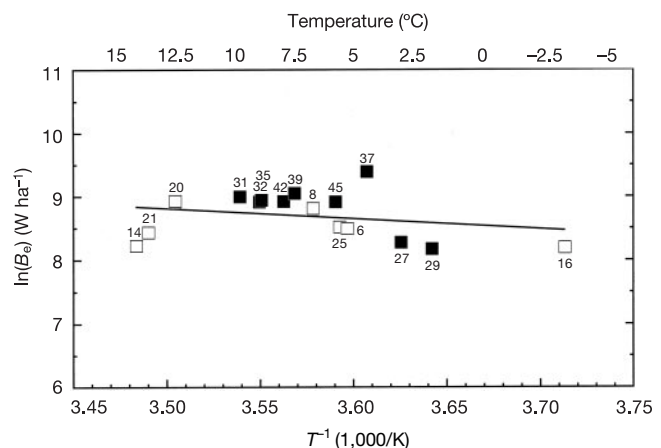


Figure 3 Relationship between the annual night-time CO_2 flux (average rate per second) and the average annual night-time temperature for both North American (open symbols) and European sites (solid symbols). Sites include 16 forests and 2 grasslands (symbols numbered 14 and 20). Note that annual night-time respiration does not vary significantly with temperature across sites (for forest ecosystems, $r^2 = 0.21$, s.e.m. = 0.337, $n = 14$, slope = -3.13 , intercept = 19.95, $P = 0.101$; see also table 1 in ref. 1). Further, as outlined by our model, ecosystem flux is not constrained by standing biomass because our grassland sites, and the agricultural site, have similar flux rates to those of the forest sites. Numbers beside each symbol refer to site numbers listed in Supplementary Table 1.

their metabolic rates (in other words, they increase b_0) to compensate for shorter growing seasons^{19–24}. The paradox illustrated here and in other recent studies^{2,9,21,22} highlights the need for further investigation of the linkages between organismal and ecosystem attributes. It also casts doubt on studies of climate change that extrapolate within-site temperature responses of organisms to regional or global scales (see refs 2, 9).

Studies of atmospheric $\delta^{13}\text{C}$ distributions strongly imply that the terrestrial biosphere contributes to inter-annual variability in global CO_2 (refs 25, 26). Our model provides a mechanistic basis for understanding how variation in resource availability, temperature and the size distribution of organisms (namely R , T and C) influence the flux of energy and materials (namely B_e , Q_{tot}) through ecological systems. The model yields many testable predictions^{10–14}. It highlights the similarity of the temperature dependence of whole-system flux, across sites comprising differing taxa, plant functional types, physiological pathways and diversity. It also explains how other prominent features of ecosystems such as standing biomass (M_e^{tot}), floristic composition and diversity might not influence ecosystem flux. Residual variation not accounted for by the model emphasizes the importance of other biotic and abiotic factors in determining ecosystem respiration (such as variability in resource availability, phenology, seasonality and life-history variation). Nevertheless, our model provides a framework for the assessment and quantitative integration of additional biotic and abiotic influences on CO_2 flux. Thus, a focus on the fundamental importance of metabolism offers a basis on which to integrate cellular, physiological and ecological attributes of ecological systems. □

Methods

We evaluated our model predictions with data from a network of micrometeorological sites that measure CO_2 , water vapour and energy exchange between the biosphere and atmosphere using the eddy covariance technique (FLUXNET⁵). FLUXNET is a unique and powerful diagnostic tool for assessing local and global scale variation in ecosystem production, respiration and net carbon exchange¹. FLUXNET consists of up to 140 sites operating on a semi-continuous basis. However, only 19 sites, with 45 site-years, provided robust continuous intra-annual measurements with multiple year records and sufficient temperature ranges. FLUXNET data have been standardized and distributed²⁷. Sites were selected from a diverse group of ecosystems, including temperate conifer and broadleaved (deciduous and evergreen) forests, boreal forest, cropland and grasslands. Together, these sites span two continents (Europe and North America) with a broad latitudinal range ($\sim 65^\circ\text{N}$ to 30°N). We also added two (non-FLUXNET) low-productivity sites from the Sonoran and Mojave Desert operated by the authors of this study. These sites each consist of 5 months' worth of continuous flux during the winter and spring seasons in 2002.

The eddy covariance technique measures net carbon and water fluxes between the biosphere and atmosphere from relatively large areas (longitudinal lengths of footprints between 100 and 2,000 m) with minimal disturbance to the system, for short and long timescales (hours, days, seasons and years; see ref. 28). Measurements of net ecosystem CO_2 exchange (NEE) during daylight hours are a function of plant photosynthesis along with both autotrophic and heterotrophic respiration. It is currently not possible to disentangle rates of respiration and photosynthesis continuously during the day (see ref. 5), but nightly values of CO_2 exchange approximate a biological ecosystem flux, or total respiration. Nightly flux values therefore provide a unique measurement of whole-system metabolism or energy flux. Here, night time is defined as the time during which the measured photosynthetically active radiation at each site is zero. To standardize ecosystem respiration, B_e , in units of metabolic energy we then converted nightly values of CO_2 flux into energy flux (W h^{-1}). Because ATP and CO_2 are both end products of respiration, the conversion is accomplished with their stoichiometry and an estimated value of available energy in quantities of ATP ($51 \text{ kJ mol}^{-1} \text{ ATP} \times (32 \text{ mol ATP/6 mol } \text{CO}_2) = 0.272 \text{ J } \mu\text{mol}^{-1} \text{ CO}_2$ (ref. 29). FLUXNET reports NEE in units of $\mu\text{mol m}^{-2} \text{ s}^{-1}$, where $(\text{NEE}) \times (0.272 \text{ J } \mu\text{mol}^{-1} \text{ CO}_2) \times (10,000 \text{ m}^2 \text{ ha}^{-1}) = B_e$ in $\text{J h}^{-1} \text{ ha}^{-1} \text{ s}^{-1}$. To parameterize equation (3) we used empirical values of biological activation energies, which vary between 0.2 and 1.2 eV, with an average of $\sim 0.6 \text{ eV}$ (refs 12, 30, 31). Transformed values of ecosystem flux were analysed with Model I regression analysis, because significant measurement error is likely to exist primarily on measures of CO_2 flux. Statistical analysis of variation in the fitted flux intercepts and slopes across sites were computed in S+ by using ANCOVA.

Studies of activation energies, E , for the respiratory complex across differing plant and animal species show little variation^{12,13,19}. Substituting the average biological activation energies predicts that the slope, $-E/1,000k$, where k is Boltzmann's constant, of the ecosystem respiration function should have a universal value of about -7.4 (ref. 12) but can range between -2 and -11 . Further, calculated activation energies for whole-organismal metabolism for both animals and plants¹² reveal a similar value of -7.79 (95% confidence interval -6.16 to -9.42).

Received 23 September 2002; accepted 20 March 2003; doi:10.1038/nature01671.

- Valentini, R. *et al.* Respiration as the main determinant of carbon balance in European forests. *Nature* **404**, 861–865 (2000).
- Giardina, C. P. & Ryan, M. G. Evidence that decomposition rates of organic carbon in mineral soil do not vary with temperature. *Nature* **404**, 858–861 (2000).
- Huxman, T. E. *et al.* Temperature as a control over ecosystem CO_2 fluxes in a high elevation subalpine forest. *Oecologia* **134**, 537–546 (2003).
- Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, 1995).
- Baldocchi, D. *et al.* FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bull. Am. Meteorol. Soc.* **82**, 2415–2434 (2001).
- Law, B. E. Environmental controls over carbon dioxide and water vapor exchange of terrestrial vegetation. *Agric. Forest Meteorol.* **113**, 97–120 (2002).
- Running, S. W. *et al.* A global terrestrial monitoring network, scaling tower fluxes with ecosystem modeling and EOS satellite data. *Remote Sens. Environ.* **70**, 108–127 (1999).
- Schulze, E. D., Kelliher, F. M., Körner, C., Lloyd, J. & Leuning, R. Relationships among maximal stomatal conductance, carbon assimilation rate, and plant nitrogen nutrition: A global ecology scaling exercise. *Annu. Rev. Ecol. Syst.* **25**, 629–660 (1994).
- Luo, Y., Wan, S., Hui, W. & Wallace, L. L. Acclimatization of soil respiration to warming in a tall grass prairie. *Nature* **413**, 622–625 (2001).
- Enquist, B. J. & Niklas, K. J. Invariant scaling relations across tree-dominated communities. *Nature* **410**, 655–660 (2001).
- West, G. B., Brown, J. H. & Enquist, B. J. A general model for the origin of allometric scaling laws in biology. *Science* **276**, 122–126 (1997).
- Gillooly, J. F., Brown, J. H., West, G. B., Savage, V. M. & Charnov, E. L. Effects of size and temperature on metabolic rate. *Science* **293**, 2248–2251 (2001).
- Gillooly, J. F. *et al.* Effects of size and temperature on developmental time. *Nature* **417**, 70–73 (2002).
- Niklas, K. J. & Enquist, B. J. Invariant scaling relationships for interspecific plant biomass production rates and body size. *Proc. Natl Acad. Sci. USA* **98**, 2922–2927 (2001).
- Hemmingsen, A. M. Energy metabolism as related to body size and respiratory surfaces, and its evolution. *Rep. Steno Mem. Hosp. Nord. Insulin Lab.* **9**, 7–95 (1960).
- West, G. B., Woodruff, W. H. & Brown, J. H. Allometric scaling of metabolic rate from molecules and mitochondria to cells and mammals. *Proc. Natl Acad. Sci. USA* **99** (suppl. 1), 2473–2478 (2002).
- Damuth, J. Population density and body size in mammals. *Nature* **290**, 699–700 (1981).
- Enquist, B. J. in *Macroecology: Concepts and Consequences* (eds Blackburn, T. & Gaston, K.) (Oxford Univ. Press, in the press).
- Raison, J. K. & Chapman, E. A. Membrane phase changes in chilling-sensitivity *Vigna radiata* and their significance to growth. *Aust. J. Plant Physiol.* **3**, 291–299 (1976).
- Pearcy, R. W. Acclimation of photosynthetic and respiratory carbon dioxide exchange to growth temperature and *Atriplex lentiformis* (Torr.) Wats. *Plant Physiol.* **59**, 795–799 (1977).
- Conover, D. O. & Schultz, E. T. Phenotypic similarity and the evolutionary significance of countergradient selection. *Trends Ecol. Evol.* **10**, 248–252 (1995).
- Mooney, H. A. & Billings, W. D. Comparative physiological ecology of arctic and alpine populations of *Oxyria digyna*. *Ecol. Monogr.* **31**, 1–29 (1961).
- Lovegrove, B. G. The zoogeography of mammalian basal metabolic rate. *Am. Nat.* **156**, 201–219 (2000).
- Jordan, C. F. A world pattern of plant energetics. *Am. Sci.* **59**, 425–433 (1971).
- Ciais, P., Tans, P. P., Trolier, M., White, J. W. C. & Francy, R. J. A large North Hemisphere terrestrial CO_2 sink indicated by the $^{13}\text{C}/^{12}\text{C}$ ratio of atmospheric CO_2 . *Science* **269**, 1098–1102 (1995).
- Keeling, C. D. & Whorf, T. P. *Trends '93: A Compendium of Data on Global Change* Oak Ridge National Laboratory Report ORNL/CDIAC-65 16–26, (1994).
- Falge, *et al.* Gap filling strategies for defensible annual sums of net ecosystem exchange. *Agric. Forest Meteorol.* **107**, 43–69 (2001).
- Baldocchi, D. D., Hicks, B. B. & Meyers, T. P. Measuring biosphere-atmosphere exchanges of biologically related gases with micrometeorological methods. *Ecology* **69**, 1331–1340 (1988).
- Nelson, D. L. & Cox, M. M. *Lehninger Principles of Biochemistry* 3rd edn (Worth Publishing, New York, 2000).
- Vetter, R. A. H. Ecophysiological studies on citrate synthase. 1. Enzyme regulation of selected crustaceans with regard to temperature adaptation. *J. Comp. Physiol. B* **165**, 46–55 (1995).
- Raven, J. A. & Geider, R. J. Temperature and algal growth. *New Phytol.* **110**, 441–461 (1988).
- Nobel, P. S. *Physicochemical and Environmental Plant Physiology* (Academic, New York, 1999).
- Earnshaw, M. J. Arrhenius plots of root respiration in some arctic plants. *Arctic Alpine Res.* **13**, 425–430 (1981).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D. Baldocchi for assistance with accessing FLUXNET data; the contributors of FLUXNET for their support of this project; and M. Weiser, D. Kerkhoff, N. Phillips, J. Harte, K. J. Niklas, S. Cowling, J. Williams and J. H. Brown for providing assistance and/or comments on earlier drafts. B.J.E. was supported by an NSF CAREER fellowship and a Center for Applied Biodiversity, Conservation International Fellowship, a LANL grant, and the University of Arizona. T.E.H. was supported by an award from IALC and the University of Arizona. B.J.E. and T.E.H. acknowledge support from the Institute for the Study of Planet Earth at the University of Arizona. J.F.G. acknowledges support from the Thaw Charitable Trust and the Packard Foundation. A.P.A. acknowledges support from NSF.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.J.E. (benquist@u.arizona.edu).

Single synaptic vesicles fusing transiently and successively without loss of identity

A. M. Aravanis^{*†‡}, J. L. Pyle^{*†} & R. W. Tsien^{*}

^{*} Department of Molecular and Cellular Physiology, Beckman Center, Stanford University School of Medicine, Stanford, California 94305, USA

[‡] Department of Electrical Engineering, Stanford University, Stanford, California 94305, USA

[†] These authors contributed equally to this work

Vesicle fusion and recycling are particularly critical for ongoing neurotransmitter release^{1–4} in the small nerve terminals of the brain, which typically contain about 30 functional vesicles^{4,5}. However, the modes of exocytosis and endocytosis that operate at synapses of the central nervous system are incompletely understood. Here we show real-time visualization of a single vesicle fusing at a small synapse of the central nervous system, made possible by highly intensified charge-coupled device imaging of hippocampal synaptic terminals, in which a single vesicle was labelled with the fluorescent membrane marker FM1-43 (ref. 6). In a small number of cases, full loss of fluorescent membrane dye was elicited by a single action potential, consistent with classical complete collapse¹. In most cases, however, action potentials triggered only partial loss of fluorescence, suggesting vesicular retention of membrane marker, consistent with 'kiss-and-run' vesicle cycling^{3,4,7–9}. An alternative hypothesis of independent fusion of partially stained vesicles arising from endosomal splitting could be excluded by observations on the size and timing of successive fusion events. Thus, our experimental evidence supports a predominance of kiss-and-run fusion events^{10–12} and rapid vesicular re-use¹¹.

We lightly labelled synaptic terminals with the styryl dye FM1-43 in order to study the real-time behaviour of single vesicles. For minimal staining, cultured hippocampal neurons were briefly exposed to dye while five stimuli were applied (Fig. 1a, first panel). After washing, complete dye release from vesicles was elicited with 1,600 stimuli. The second and third panels of Fig. 1a show fluorescence images of a stained area before and after stimulation for de-staining. After a brief rest, all functional presynaptic boutons in the same fields were identified by high K⁺ depolarization in the presence of FM1-43 (maximal staining protocol; Fig. 1b, first panel). Images of fully loaded terminals (Fig. 1b, second panel) defined regions of interest (for example, ROIs 1–3) for retrospective analysis of dye uptake and release resulting from minimal staining (Fig. 1a, second panel). In this field, only bouton 2 took up and released fluorescent membrane dye during minimal staining (Fig. 1a, second and fourth panels); boutons 1 and 3 remained at background levels throughout.

The stimulus-induced fluorescence loss ($\Delta F_{\max}$) in 272 minimally stained boutons is displayed in Fig. 1c in units of fluorescein-isothiocyanate (FITC) molecular equivalents. The histogram shows discrete peaks reflecting the quantal nature of fluorescent membrane dye staining^{4,13,14}. The distribution was well-fitted by the sum of individual gaussians conforming to expectations of quantal theory¹⁵: the 0th peak represents functional synapses that failed to take up FM1-43; the first and second peaks correspond to terminals loaded with one and two vesicles, respectively; the spacing between successive peaks is the average fluorescence of single vesicles ($\bar{q}$). Here, $\bar{q}$ was 500 FITC equivalents. To verify that the fluorescence quantum originates from single vesicles, we labelled neuronal surface membrane with FM1-43 and determined that 500 FITC equivalents was the fluorescence of 1,684 nm² of membrane, an

area closely matching that of vesicles (inner diameter approximately 24 nm, N. Harata, unpublished data). As a further check, we took the average fluorescence of maximally loaded single boutons ($14,450 \pm 450$ FITC equivalents, $n = 272$, Fig. 1d; mean $\pm$ standard error of the mean (s.e.m.) here and elsewhere) and divided by 30 vesicles, which is the total number of recycling vesicles estimated using FM1-43 photoconversion and electron microscopy in our hippocampal culture system⁴. This yielded a morphologically based estimate of $\bar{q}$ at 482 FITC equivalents. Thus, several independent lines of evidence verified that $\bar{q}$ represents the fluorescence of single dye-labelled vesicles.

Figure 2a shows the real-time response to 10 Hz stimulation of individual boutons containing zero, one, or two dye-stained vesicles. Hereafter, we focused our analysis on boutons containing one dye-stained vesicle with >95% certainty (shaded area of histogram, Fig. 2a). Their fluorescence signals were analysed for time-to-onset of first exocytosis ($t_{\text{onset}} \equiv t_{20\%}$) and the time for near-complete quantal loss ($t_{20-80\%}$) (Fig. 2b, c). In 65% of cases (Fig. 2d), t_{onset} was ≤ 4 s (40 action potentials), indicating that the single vesicle was stationed in the readily releasable pool (RRP)^{5,16}. The RRP comprises only one-quarter to one-third of the total recycling pool in these terminals^{5,11}, so less than one-third of the single vesicles would have resided in the RRP had they mixed randomly in the terminal. The readiness of these single vesicles implies that, after minimal stimulation, random mixing did not occur, as it does with more intense stimulation^{5,11}.

Although t_{onset} was brief, the de-staining time of a single vesicle ($t_{20-80\%}$) was long, averaging 16.9 ± 2.3 s ($n = 20$) (Fig. 2e). This was not compatible with a model of exocytosis in which fusion always leads to complete collapse—in that case, $t_{20-80\%}$ would be < 3.5 s (ref. 10). This disparity was highlighted by averaging records of single-vesicle de-staining, time-aligned at t_{onset} (Fig. 2f). The average signal was best fitted with a single exponential with $\tau = 10.3$ s, considerably slower than $\tau_{\text{departition}} = 2.5$ s predicted for complete collapse with aqueous departitioning¹⁰, and much slower than the $\tau_{\text{lateral}} \approx 10$ ms predicted for lateral diffusion¹⁷ (smooth curves, Fig. 2f). This suggests that, in general, vesicles did not completely collapse but underwent a rapid retrieval that supported dye retention for several seconds. The full de-staining eventually observed with additional stimulation implied that single vesicles fused more than once¹¹.

Next, we monitored the exocytosis of single vesicles after single action potentials (Fig. 3). The stimulation frequency was reduced to 0.125 Hz to allow enough time between action potentials for complete dye re-equilibration after individual fusion events. After 20 stimuli (80 images), imaging was interrupted, and 600 stimuli at 10 Hz were applied to completely de-stain the vesicle (Fig. 3a). The quantal size in these experiments was 488 FITC equivalents (Fig. 3f), similar to Fig. 1c. Notably, some single vesicles released a full quantum of dye in response to one action potential (Fig. 3b), seen as an immediate decrease in fluorescence amounting to about 100% of the quantal amplitude (Fig. 3c). Near-complete loss of fluorescence in one action potential was consistent with complete collapse of the vesicle into the plasma membrane^{1,17}, although we cannot exclude the possibility that the fusion pore was transient but remained patent long enough for all dye to escape.

Recordings such as Fig. 3c were rare ($n = 3$). In most cases ($n = 17$) vesicle fluorescence dropped in response to at least one of the 20 action potentials, but not to the extent of a full quantum (Fig. 3d, e). This subquantal loss of dye also followed a single action potential, and presumably arose from a fusion event that did not result in complete collapse. Even after 20 action potentials, the fluorescence loss from single vesicles averaged only $0.67\bar{q}$ ($n = 20$, Fig. 3g). The simplest interpretation is that most fusion events were followed by rapid retrieval that allowed partial dye retention. The asymmetry between loading, which is quantal, and unloading, which is mostly subquantal, can be explained by the rapid rate of

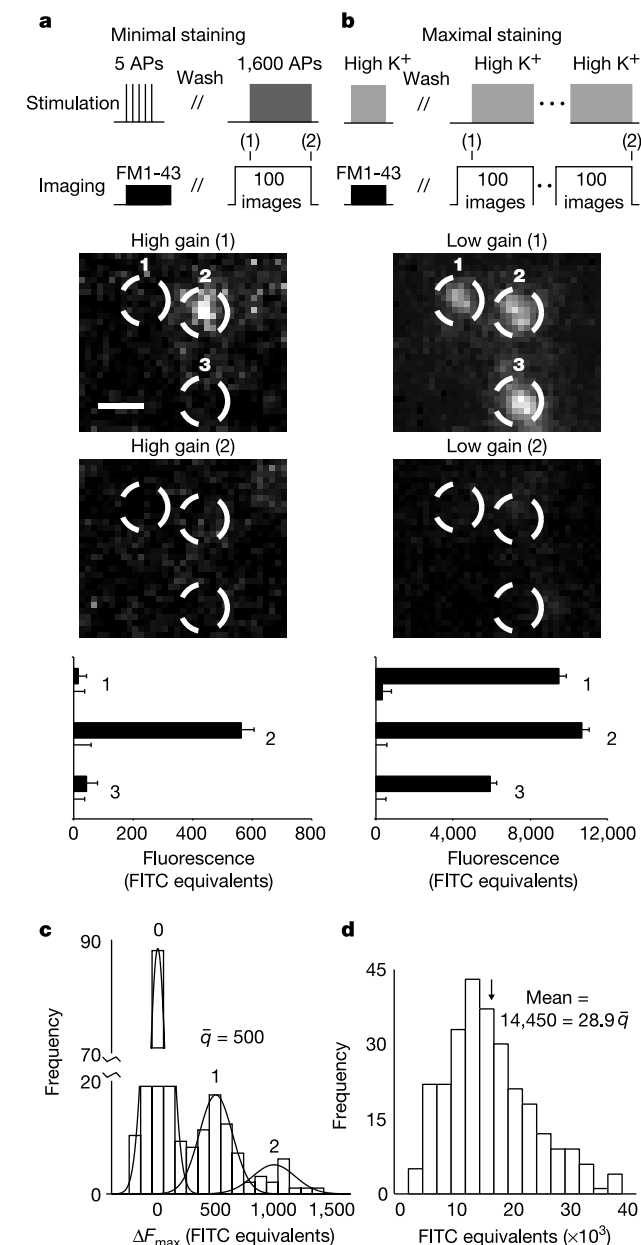


Figure 1 Quantal staining of FM1-43 at single boutons. **a**, Minimal-staining protocol: field stimulation with five stimuli at 10 Hz in 16 μ M FM1-43 (first panel). The second panel shows a minimally stained region. ROIs 1, 2, and 3 are centred on functional presynaptic terminals (see **b**, second panel). Bouton 2 contained one synaptic vesicle labelled with FM1-43; boutons 1 and 3 contained none. Scale bar, 1 μ m. The third panel shows the same area as the second but after 1,600 stimuli (10 Hz). The fourth panel shows fluorescence intensities of each ROI before (top bar) and after (bottom bars) stimulation, as in the second and third panels. Error bars indicate standard deviation of 20 images. AP, action potential. **b**, Maximal-staining protocol (first panel). The second panel is the same area as the second panel of **a** after maximal staining. The third panel is the same area as the second panel after exhaustive de-staining. The fourth panel shows fluorescence intensities of the ROIs before (top bar) and after (bottom bars) stimulation, as in the second and third panels. **c**, Histogram of the fluorescence change ($\Delta F_{\max}$) of minimally stained terminals ($n = 272$). Peaks represent boutons containing zero, one, or two vesicles; quantal spacing ($\bar{q}$), 500 FITC equivalents. **d**, Histogram of the number of functional synaptic vesicles in CA1–CA3 terminals.

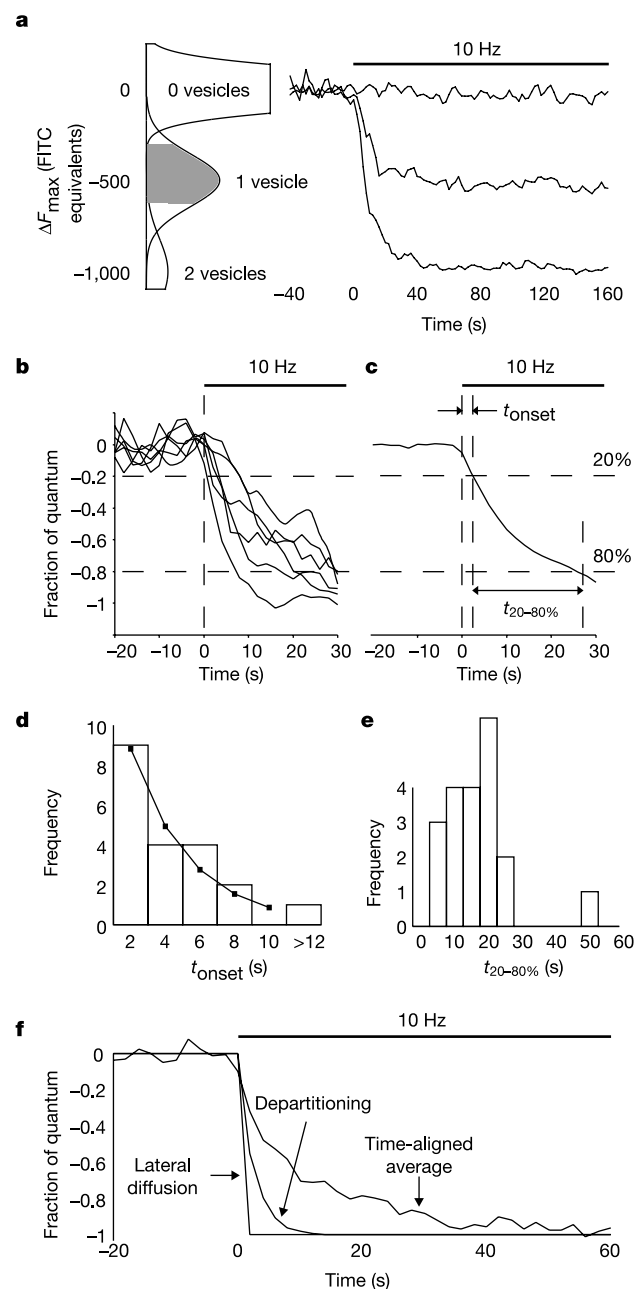


Figure 2 Single vesicles loaded during minimal stimulation remain in the RRP and retain dye after first fusion. **a**, Fluorescence signals from individual boutons containing 0, 1, or 2 vesicles during 10 Hz stimulation. The shaded area in the histogram indicates single vesicles with >95% confidence. **b**, Time-expanded records of single-vesicle de-staining. **c**, Average of single-vesicle de-staining ($n = 20$) with analysis scheme. t_{onset} , time from stimulus initiation to >20% dye loss; $t_{20-80\%}$, time from >20% to >80% dye loss. **d**, Histogram of t_{onset} (4.6 ± 1.0 s, $n = 20$). The solid line is a fitted geometric distribution with probability of release per vesicle (P_{rv} at 10 Hz) of 0.03. **e**, Histogram of $t_{20-80\%}$. **f**, Average of time-aligned traces ($n = 20$; normalized, aligned to t_{onset} and averaged). A single exponential fit gave $\tau = 10.3$ s, much slower than time courses of FM1-43 release predicted by ideal dye departioning or ideal lateral diffusion.

dye partitioning from solution into membrane relative to the much slower dissociation rate^{6,18}.

We analysed the timing and magnitude of dye loss in response to single action potentials to gain information on individual fusion events (Fig. 4). Figure 4a shows the fluorescence signal from a vesicle that fused twice within the first 20 stimuli. Both fusion events were well resolved, each releasing >100 FITC equivalents. Such single vesicle traces were fitted with a staircase

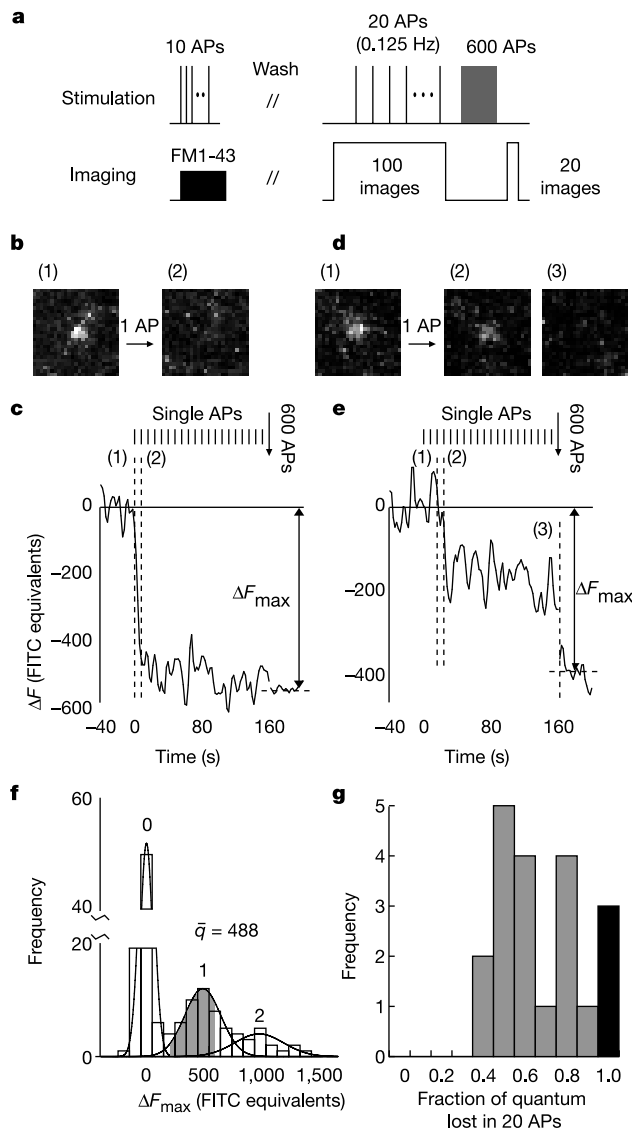


Figure 3 Individual fusion events usually result in dye retention. **a**, Low-frequency stimulation (4:1 ratio of imaging to action potentials) allowed the consequences of individual action potentials to be resolved. **b**, Individual high-gain images, taken just before the first (left) and second (right) stimuli. A single vesicle's worth of fluorescence was lost after one action potential was elicited. **c**, Corresponding fluorescence signal. **d**, Individual high-gain images from another experiment, taken just before the third (left) and fourth (middle) stimuli and just after the 620th (right) stimuli. A fusion event after the third action potential was followed by retrieval rapid enough to leave FM1-43 trapped in the vesicle. The remaining fluorescence was retained for 17 action potentials and then released with 600 action potentials. **e**, Corresponding fluorescence signal. **f**, Quantal analysis of fluorescence change before and after exhaustive stimulation ($n = 153$ ROIs). The shaded area in the histogram represents single stained vesicles (>95% confidence) that were taken for analysis. **g**, Most single vesicles (85%) retained a fraction of their original quantum of dye after 20 action potentials, whereas 15% lost the full quantum in one action potential (black bar).

function to determine the timing and magnitude of FM1-43 loss. If one vesicle were responsible for both fusions and the second fusion event were similar to the first, one would expect the absolute amount of dye loss in the second fusion to be less than in the first, and a similar fraction of the remaining dye to escape in both fusions. Indeed, the amount of dye lost in the second fusion (0.30 q) was significantly less on average than the first (0.48 q) ($P < 0.05$, t -test, Fig. 4b); the fractional dye loss in second fusions (0.42 q' , where q' is the fluorescence left after previous events) was not significantly different from first fusions ($P > 0.5$, t -test, Fig. 4c). These data also set an upper bound on the lifetime of the fusion event of <1.4 s (for aqueous departitioning¹⁰).

Analysis of the latency to first fusion of single synaptic vesicles showed that on average, dye was first released after only 4.9 stimuli ($n = 19$, Fig. 4d), once again suggesting that these vesicles were ready for immediate use. In traces where a second fusion event was resolved within 20 action potentials (for example, Fig. 4a), we determined the interval between the first and second fusions (Fig. 4d). This second-fusion latency averaged 7.1 stimuli. First event latency was roughly exponentially distributed, conforming to a geometric function (Fig. 4d, black smooth curve) with a release probability per stimulus ($P_{r/v}$ at 0.125 Hz) of 0.20. The distribution of second-fusion latency was significantly different ($P < 0.01$, Kolmogorov–Smirnov test), being distinguished by a sigmoid onset. To model this, we assumed that the second release obeyed kinetics similar to the first, but only after vesicle re-priming, described with an exponentially distributed interval ($\tau_{\text{re-prime}}$ at 0.125 Hz). In the best fit (Fig. 4d, grey smooth curve), $\tau_{\text{re-prime}}$ at 0.125 Hz was 23 s. A significant extra delay before second fusion was also apparent when single-vesicle de-staining records were time-aligned and averaged (Fig. 4e). After the initial rapid drop, this average response levelled off for a few stimulus intervals before undergoing a secondary increase in slope.

We considered an alternative interpretation of these findings, in which a single dye-stained vesicle buds from the plasma membrane but somehow gives rise to two or more partly stained synaptic vesicles that then fuse independently. Several lines of evidence render this possibility unlikely, if not untenable. First, this model conflicts with the findings that second fusions release a smaller absolute amount of dye (Fig. 4b) but a similar fraction of remaining dye (Fig. 4c). Second, the extra delay between the initial fluorescence step and a second step (Fig. 4d, e) contradicts expectations for independent fusions. Third, even after dye loading with strong stimulation, fluorescent membrane photoconversion and electron microscopy indicate that endosomal structures in hippocampal terminals are extremely rare (N. Harata, unpublished data).

Improvements in instrumentation and experimental protocols enabled us to track the real-time de-staining of single fluorescent membrane-labelled vesicles in small terminals of the central nervous system. Most of the single-vesicle fusion events were marked by only partial loss of dye. This was seen as a subquantal drop in fluorescence after individual action potentials at low frequency and as a slow de-staining time course at high frequency. A likely explanation is that most fusion events were simply too brief to permit dye to escape completely¹⁰ (compare with ref. 17). After rapid retrieval by a process that maintained vesicle integrity^{8,19,20}, vesicles remained available for repeated fusion (Fig. 4), supporting their re-use for neurotransmitter release^{11,12}. Together, brief fusion, rapid retrieval and re-use would enable small nerve terminals of the central nervous system to get full mileage from their limited set of vesicles⁴. Re-use of RRP vesicles during routine neurotransmission would minimize recruitment of reserve pool vesicles until they were absolutely required for prolonged high-frequency transmission. With all-out vesicular mobilization, the prevalent mode of fusion/retrieval may shift in favour of classical exo/endocytosis¹ and *de novo* re-creation of vesicles^{11,21,22}. A combination of multiple fusion and retrieval modes would be consistent with our data, and may

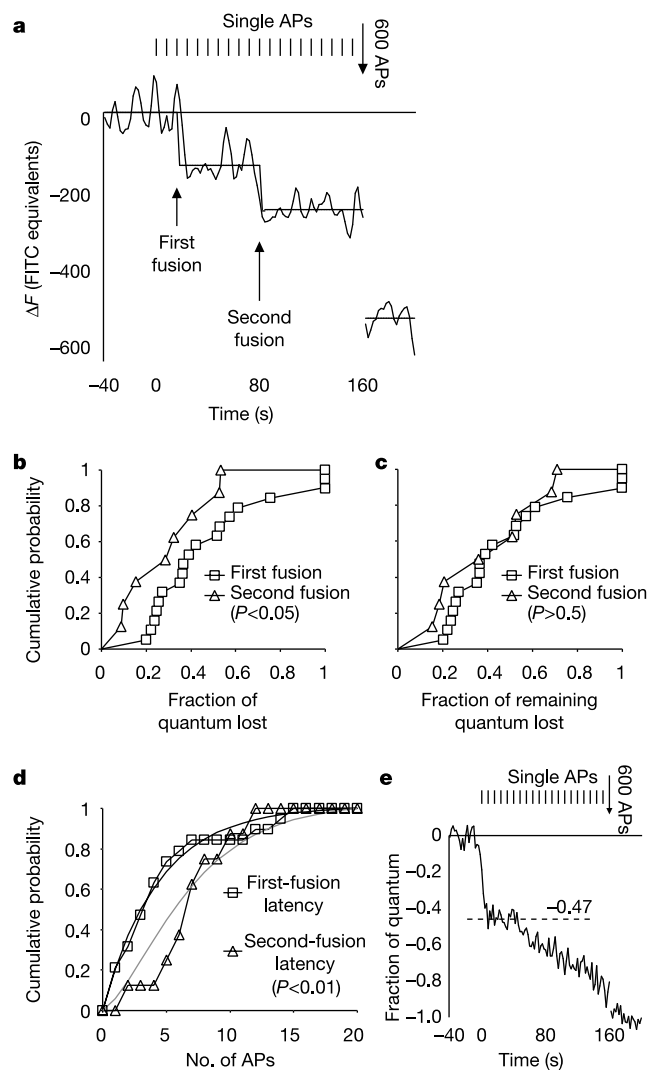


Figure 4 Single vesicles can fuse more than once during 20 action potentials. **a**, A vesicle fused twice (third and eleventh action potentials). **b**, Collected data of fluorescence lost on first ($n = 19$) and second fusions ($n = 8$), shown as cumulative distributions. Mean ΔF on first fusions, $0.48q$, and on second fusions, $0.30q$ ($P < 0.05$, t -test). **c**, Fraction of remaining quantum lost on first and second fusions, given as cumulative distributions (first mean = $0.48q$, second mean = $0.42q$; $P > 0.5$, t -test). **d**, Cumulative distributions of first-fusion latency (mean = 4.9 action potentials, $n = 19$) and second-fusion latency (interval between first and second fusions; mean = 7.1 action potentials, $n = 8$). First-fusion latency was significantly faster than second-fusion latency ($P < 0.01$, Kolmogorov–Smirnov test), indicating that another process delays the second fusion. **e**, Average of time-aligned traces ($n = 19$), emphasizing that the first fusion resulted in loss of 47% of the quantum on average, and that second fusion was significantly delayed.

be well suited for patterns of firing found in the hippocampus *in vivo*. □

Methods

Cell cultures

Rat hippocampal CA3–CA1 neurons were prepared in sparse culture²³, and used after 12–18 days. To achieve the stringent conditions necessary for imaging, >600 individual coverslips (50 independent cultures) were examined to obtain enough suitable samples (10 Hz, six coverslips; 0.125 Hz, eight coverslips). We performed all experiments at 25 °C, in physiological solutions containing 2 mM Ca^{2+} and 2 mM Mg^{2+} .

Dye loading and de-staining

Presynaptic terminals were labelled by exposure to 16 μM FM1-43 while 5 (Figs 1 and 2) or 10 (Figs 3 and 4) field stimuli were delivered at 10 Hz (platinum bath electrodes, 20 mA, 1-ms pulses). After 10 s of dye exposure to allow full staining of single vesicles, dye was

removed by washing for 15 min with dye-free Tyrode solution (containing no Ca^{2+} to minimize spontaneous exocytosis). All solutions contained 10 μM 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX) to prevent recurrent activity. Hippocampal terminals were de-stained by additional field stimulation at 10 Hz (1,600 action potentials) or at 0.125 Hz (20 action potentials) followed by 600 action potentials at 10 Hz. After 3–4 min recovery, the entire recycling pool of all synapses was stained by a 90-s exposure to 16 μM FM1-43 in 45 mM K^{+} solution⁴. Images of the maximally stained field were acquired before and after complete de-staining with three rounds of high K^{+} depolarization (60 s each).

Imaging

Fluorescence detection of single synaptic vesicles was performed with an inverted epifluorescence microscope (1.3 NA objective) and an intensified charge-coupled device camera operating in gated acquisition mode (XR/Mega-10, Stanford Photonics). Brief pulses of illumination (470/40 nm) were gated by an optical switch; fluorescence emission passed through a 515-nm long-pass filter. The exposure time per image (15 ms) helped maximize fluorescence signals while causing acceptably mild photobleaching (<10% over 121 images). In the maximal-staining protocol, intensifier gain and exposure time were lowered to decrease overall system gain by a factor of 15. Images were acquired at 0.5 Hz. Stimulation was phase-locked to the imaging, with the first action potential occurring milliseconds after the 21st image. ΔF_{max} was the difference between 20-image averages before and after stimulation.

Calibration

Fluorescence intensities were calibrated by imaging 0.87- μm beads coated with 980 ± 15 FITC molecules (Bangs Laboratories). The imaging system point spread function (p.s.f.) was determined empirically by repeated imaging of 49-nm beads coated with fluorescent molecules with FM1-43-like emission. A 1.09- μm circle (eight pixels), corresponding to 6.8 standard deviations of the fitted p.s.f. gaussian, was chosen as the standard ROI. With this ROI size, and a 0.7–1.0- μm depth of field, the system was tolerant to variations in vesicle position within typical hippocampal terminals (diameter <1 μm). In fact, changes in vesicle position were uncommon.

Identification of individual synapses

Synapses were identified using maximal staining protocol through uptake of FM1-43 during 90-s application of 45 mM K^{+} , and accepted for analysis if they appeared as punctate regions <1.2 μm in diameter, showing a ΔF signal >2,500 FITC equivalents (five vesicles). A circular ROI was centred on the intensity peak.

Selection of single vesicles

Histograms of minimal-staining intensities were fitted with the sum of three gaussians:

$$T(\Delta F_{\text{max}}) = \sum_{k=0}^2 v_k e^{-\frac{(\Delta F_{\text{max}} - kq)^2}{2(\sigma_0^2 + kq^2)}} \quad (1)$$

where T is the number of events at a given ΔF_{max} , v_k is the amplitude of the k th peak, σ_0^2 is the variance of the 0th peak (measurement error), σ_k^2 is the variance of the fluorescence of a single FM1-43-labelled vesicle, and q is the quantal fluorescence. The range of ΔF_{max} values corresponding to single vesicles was estimated by requiring a >95% probability of belonging to the $k = 1$ gaussian.

Analysis of individual fusion events

Fusion events were defined as abrupt and lasting fluorescent decreases >10% of ΔF_{max} , which represented true signal changes with >95% confidence. A staircase with the same number of events was best fitted to the trace, with timing and magnitude of steps as free parameters.

Received 9 March; accepted 28 April 2003; doi:10.1038/nature01686.

- Heuser, J. E. & Reese, T. S. Evidence for recycling of synaptic vesicle membrane during transmitter release at the frog neuromuscular junction. *J. Cell Biol.* **57**, 315–344 (1973).
- Cremona, O. & De Camilli, P. Synaptic vesicle endocytosis. *Curr. Opin. Neurobiol.* **7**, 323–330 (1997).
- Sudhof, T. C. The synaptic vesicle cycle revisited. *Neuron* **28**, 317–320 (2000).
- Harata, N. *et al.* Limited numbers of recycling vesicles in small CNS nerve terminals: implications for neural signaling and vesicular cycling. *Trends Neurosci.* **24**, 637–643 (2001).
- Murthy, V. N. & Stevens, C. F. Reversal of synaptic vesicle docking at central synapses. *Nature Neurosci.* **2**, 503–507 (1999).
- Cochilla, A. J., Angleson, J. K. & Betz, W. J. Monitoring secretory membrane with FM1-43 fluorescence. *Annu. Rev. Neurosci.* **22**, 1–10 (1999).
- Ceccarelli, B., Hurlbut, W. P. & Mauro, A. Turnover of transmitter and synaptic vesicles at the frog neuromuscular junction. *J. Cell Biol.* **57**, 499–524 (1973).
- Palfrey, H. C. & Artalejo, C. R. Vesicle recycling revisited: rapid endocytosis may be the first step. *Neuroscience* **83**, 969–989 (1998).
- Stevens, C. F. & Williams, J. H. 'Kiss and run' exocytosis at hippocampal synapses. *Proc. Natl Acad. Sci. USA* **97**, 12828–12833 (2000).
- Klingauf, J., Kavalali, E. T. & Tsien, R. W. Kinetics and regulation of fast endocytosis at hippocampal synapses. *Nature* **394**, 581–585 (1998).
- Pyle, J. L., Kavalali, E. T., Piedras-Renteria, E. S. & Tsien, R. W. Rapid reuse of readily releasable pool vesicles at hippocampal synapses. *Neuron* **28**, 221–231 (2000).
- Sara, Y., Mozhayeva, M. G., Liu, X. & Kavalali, E. T. Fast vesicle recycling supports neurotransmission during sustained stimulation at hippocampal synapses. *J. Neurosci.* **22**, 1608–1617 (2002).
- Ryan, T. A., Reuter, H. & Smith, S. J. Optical detection of a quantal presynaptic membrane turnover. *Nature* **388**, 478–482 (1997).
- Murthy, V. N. & Stevens, C. F. Synaptic vesicles retain their identity through the endocytic cycle. *Nature* **392**, 497–501 (1998).

15. Del Castillo, J. & Katz, B. Quantal components of the end plate potential. *J. Physiol. (Lond.)*, **124**, 560–573 (1954).
16. Rosenmund, C. & Stevens, C. F. Definition of the readily releasable pool of vesicles at hippocampal synapses. *Neuron* **16**, 1197–1207 (1996).
17. Zenisek, D., Steyer, J. A., Feldman, M. E. & Almers, W. A membrane marker leaves synaptic vesicles in milliseconds after exocytosis in retinal bipolar cells. *Neuron* **35**, 1085–1097 (2002).
18. Neves, G. & Lagnado, L. The kinetics of exocytosis and endocytosis in the synaptic terminal of goldfish retinal bipolar cells. *J. Physiol. (Lond.)* **515**, 181–202 (1999).
19. Klyachko, V. A. & Jackson, M. B. Capacitance steps and fusion pores of small and large-dense-core vesicles in nerve terminals. *Nature* **418**, 89–92 (2002).
20. Sun, J. Y., Wu, X. S. & Wu, L. G. Single and multiple vesicle fusion induce different rates of endocytosis at a central synapse. *Nature* **417**, 555–559 (2002).
21. Stevens, C. F. & Wesseling, J. F. Identification of a novel process limiting the rate of synaptic vesicle cycling at hippocampal synapses. *Neuron* **24**, 1017–1028 (1999).
22. Sankaranarayanan, S. & Ryan, T. A. Real-time measurements of vesicle-SNARE recycling in synapses of the central nervous system. *Nature Cell Biol.* **2**, 197–204 (2000).
23. Malgaroli, A. & Tsien, R. W. Glutamate-induced long-term potentiation of the frequency of miniature synaptic currents in cultured hippocampal neurons. *Nature* **357**, 134–139 (1992).

Acknowledgements We thank R. W. Aldrich and N. Harata for critically reading the manuscript; S. J. Smith for advice; and members of the Tsien laboratory for comments. This work was supported by grants from the NIMH (R.W.T.) and the Medical Scientist Training Program (A.M.A. and J.L.P.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.W.T. (rwttsien@stanford.edu).

Fission yeast *mod5p* regulates polarized growth through anchoring of *tea1p* at cell tips

Hilary A. Snaith & Kenneth E. Sawin

Wellcome Trust Centre for Cell Biology, Institute of Cell and Molecular Biology, Swann Building, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JR, UK

Microtubules have a central role in eukaryotic cell polarity¹, in part through interactions between microtubule end-binding proteins and the cell cortex^{2,3}. In the fission yeast *Schizosaccharomyces pombe*, microtubules and the polarity modulator *tea1p* maintain cylindrical cell shape and strictly antipodal cell growth^{4–7}. The *tea1p* protein is transported to cell tips by association with growing microtubule plus ends⁸; once at cell tips, *tea1p* releases from microtubule ends and associates with the cell cortex, where it coordinates polarized growth^{4,6}. Here we describe a cortical protein, *mod5p*, that regulates the dynamic behaviour of *tea1p*. In *mod5Δ* cells, *tea1p* is efficiently transported on microtubules to cell tips but fails to anchor properly at the cortex and thus fails to accumulate to normal levels. *mod5p* contains a signal for carboxy-terminal prenylation and in wild-type cells is associated with the plasma membrane at cell tips. However, in *tea1Δ* cells, although *mod5p* remains localized to the plasma membrane, *mod5p* is no longer restricted to the cell tips. We propose that *tea1p* and *mod5p* act in a positive-feedback loop in the microtubule-mediated regulation of cell polarity.

From an insertional mutagenesis screen designed to identify non-essential genes regulating cell polarity in fission yeast (H.A.S., unpublished observations), we identified four genes whose loss-of-function phenotype resembles that of a *tea1Δ* strain, which forms bent or branched cells after a variety of stresses^{4,5} (see Supplementary Information). One of these genes was *tea1⁺* itself (Fig. 1a), and two of the others were previously identified genes that are known to affect microtubule organization and consequently the normal

delivery of *tea1p* to cell tips (*tea2⁺*, *tip1⁺* (refs 9, 10)). The fourth gene, which we termed *mod5⁺* (for morphology defective 5), has not previously been characterized and is identified as open reading frame SPBC530.04 in the *S. pombe* genome (http://www.sanger.ac.uk/cgi-bin/yeastpub/pombe_chr_status). Deletion of the complete open reading frame of *mod5⁺* (*mod5Δ*) did not affect cell viability and yielded the same mutant phenotype as the original insertion mutant (Fig. 1b).

In immunofluorescence experiments with wild-type cells, *tea1p* was present both at cell tips and at the ends of microtubules (Fig. 1c,

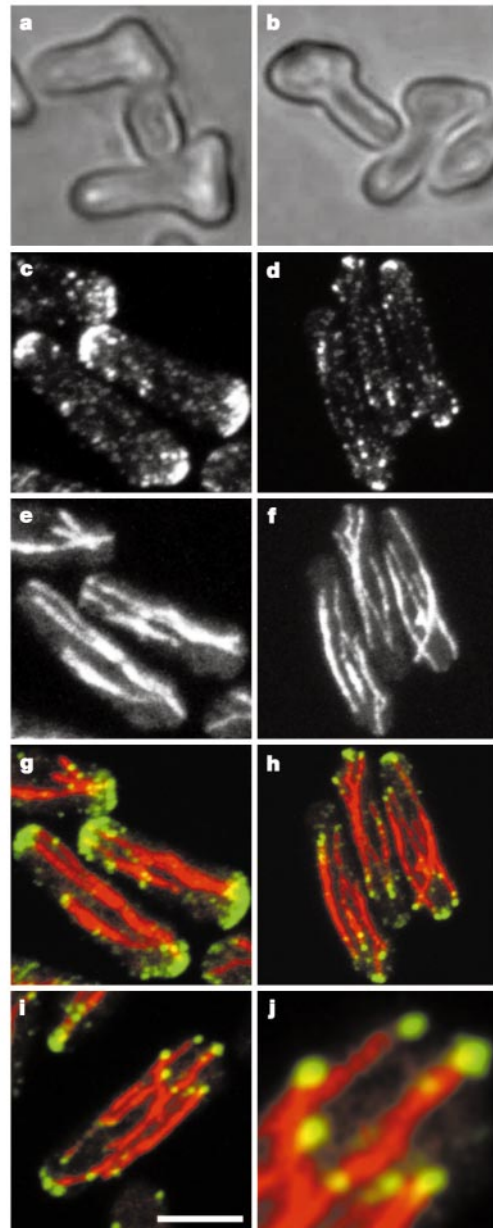


Figure 1 *mod5Δ* cells fail to localize *tea1p* at cell ends. **a, b**, Phenotype of *tea1Δ* cells (**a**) and *mod5Δ* cells (**b**) on solid medium 4 h after refeeding. Wild-type cells are uniformly cylindrical (see, for example, **c, d**). **c, d**, Anti-*tea1p* staining in wild-type cells (**c**) and *mod5Δ* cells (**d**). **e, f**, Anti-tubulin staining in wild-type cells (**e**) and *mod5Δ* cells (**f**). **g, h**, Merged images of **c** and **e**, and **d** and **f**, respectively. *tea1p* accumulates at cell tips in wild-type cells but is restricted to microtubule ends in *mod5Δ* cells. **i**, *mod5Δ* cell stained for tubulin and *tea1p*, showing *tea1p* on microtubule ends. **j**, Enlargement of the cell tip shown in **i**. Scale bar, 5 μm (**a–i**) and 1 μm (**j**).

e, g)⁴. Strikingly, in *mod5Δ* cells, although tea1p remained associated with microtubule ends, it no longer accumulated to high levels at cell tips (Fig. 1d, f, h–j; see also Supplementary Information), and 91% of tea1p spots observed at cell tips in *mod5Δ* cells were associated with a microtubule end ($n = 613$ spots). Immunoblotting experiments confirmed that tea1p levels are not altered in *mod5Δ* cells (data not shown). Microtubule organization was also generally similar between wild-type and *mod5Δ* cells (Fig. 1e, f), although a small percentage of *mod5Δ* cells contained a microtubule curling around the cell tip^{4,8} (see Supplementary Information).

Because tea1p is normally targeted to cell ends by microtubules^{4,8}, we further investigated the role of mod5p in the cortical localization of tea1p by disrupting microtubules during polarity re-establishment experiments, in which cells are first grown to stationary phase and then diluted into fresh medium. This procedure increases the penetrance of polarity mutant phenotypes during the first cell cycle after dilution into fresh medium (ref. 9 and our unpublished observations) and was performed both in the presence and in the absence of the fungal microtubule inhibitor 2-methyl benzimidazolylcarbamate (MBC). Within 3 hours of being returned to fresh medium, wild-type cells not treated with the drug re-established a growth polarity axis, with tea1p uniformly decorating cell tips (Fig. 2a). By contrast, in *mod5Δ* cells not treated with drug, tea1p was found only at microtubule ends and did not accumulate at cell tips (Fig. 2b). It has been reported that the localization of tea1p at cell tips is strongly dependent on microtubules⁴. However, in wild-type cells treated with MBC during polarity re-establishment, at

least one cell tip was able to accumulate significant levels of tea1p, despite the presence of only very short microtubule remnants (Fig. 2c). In contrast, *mod5Δ* cells exhibited nearly no detectable tea1p at cell tips after treatment with MBC (Fig. 2d; see Supplementary Information). We conclude that a major defect associated with the loss of mod5p is a failure to retain tea1p at cell tips, and that this is likely to be independent of the microtubule-based targeting of tea1p.

These results also indicate that the targeting of tea1p to cell tips might be a two-step process in which microtubule-dependent delivery is followed by a microtubule-independent, but mod5p-dependent, cortical anchoring mechanism. We speculate that, in the absence of microtubules, diffusion of tea1p might allow its eventual cortical association, provided that mod5p is present. This interpretation is supported by the frequency of abnormal, branched cells observed in polarity re-establishment experiments (Fig. 2e). In *tea1Δ* cells a high frequency of cell branching was seen both in the absence and in the presence of MBC (88% and 92% of cells, respectively; $n = 200$ cells). In comparison, very few wild-type cells formed branches under either condition (0% and 3%, respectively). However, in *mod5Δ* cells the frequency of branched cells increased markedly in the presence of MBC, from 1% to 84%. We interpret these data to indicate, first, that the absence of tea1p from cell tips during polarity re-establishment leads to a high frequency of branching; second, that in *mod5Δ* cells not treated with the drug, the small amount of tea1p present at cell tips in association with microtubule ends might often be sufficient for normal polarized growth; and last, that in MBC-treated *mod5Δ* cells, the further loss of this small amount of microtubule-associated tea1p produces a phenocopy of *tea1Δ*.

To determine whether mod5p might have a role not only in the anchoring of tea1p at the cortex but also in the microtubule-based transport of tea1p, we examined live cells expressing green fluorescent protein (GFP)-tagged derivatives of tea1p and α -tubulin (atb2p, the non-essential α -tubulin in *S. pombe*^{11–14} (K.E.S., unpublished observations)). In cells expressing GFP-tubulin at near-endogenous levels, no significant differences in microtubule dynamics were apparent between wild-type and *mod5Δ* cells (data not shown). The localization of tea1p–GFP in live wild-type and *mod5Δ* cells was similar to that of endogenous tea1p in fixed cells (Fig. 3a, b; see also Supplementary Information). Time-lapse videomicroscopy of wild-type cells revealed small particles of tea1p–GFP originating in the medial perinuclear region of the cell and translocating towards the cell tips at $2.43 \pm 0.40 \mu\text{m min}^{-1}$ (mean $\pm$ s.d.; $n = 27$ particles; Fig. 3a), where they were then retained, as described recently⁸. In *mod5Δ* cells, tea1p–GFP particles moved from the nuclear periphery towards the cell tips at similar rates ($2.69 \pm 0.55 \mu\text{m min}^{-1}$; $n = 27$ particles; Fig. 3b) but upon reaching cell tips either moved away from the cortex and/or disassembled, which is consistent with observations of fixed time-points (see Supplementary Information).

Three other differences in the properties of tea1p–GFP particles were also apparent in *mod5Δ* cells. First, *mod5Δ* cells contained 6.52 ± 0.98 tea1p–GFP particles per cell (mean $\pm$ s.d.; $n = 116$ cells), compared with 2.73 ± 1.24 in wild-type cells ($n = 131$ cells). In addition, tea1p–GFP particles in *mod5Δ* cells were 2–3-fold brighter than in wild-type cells (see Supplementary Information). Moreover, in many instances several tea1p–GFP particles seemed to move in a single stream towards the cell tips (Fig. 3b; see also Supplementary Information). The increased number and intensity of particles can be explained in part by the fact that *mod5Δ* mutants might contain a larger free cytoplasmic pool of tea1p–GFP because it is no longer sequestered at cell tips. In wild-type cells, tea1p–GFP particles translocating towards cell tips have been shown to be specifically associated with the plus ends of growing microtubule bundles⁸. The streaming behaviour of tea1p–GFP particles observed in *mod5Δ* cells indicated that several particles might be tracking

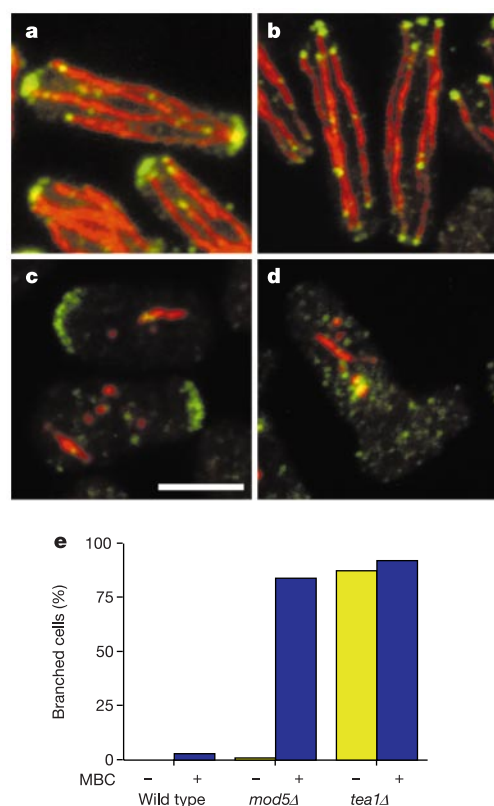


Figure 2 Microtubule-independent targeting of tea1p to the cortex depends on mod5p. Cells are stained for tea1p (green) and tubulin (red). **a, b**, Wild-type cells (**a**) and *mod5Δ* (**b**) cells 3 h after release to growth in the presence of dimethyl sulphoxide (control). **c, d**, Wild-type cells (**c**) and *mod5Δ* cells (**d**) 3 h after release to growth in the presence of $50 \mu\text{g ml}^{-1}$ MBC. Scale bar, 5 μm . **e**, Percentage of cells forming branches in wild-type, *mod5Δ* and *tea1Δ* cells 3 h after release to growth, in the absence and in the presence of MBC. The percentages shown in the first three columns are 0%, 3% and 1%, respectively.

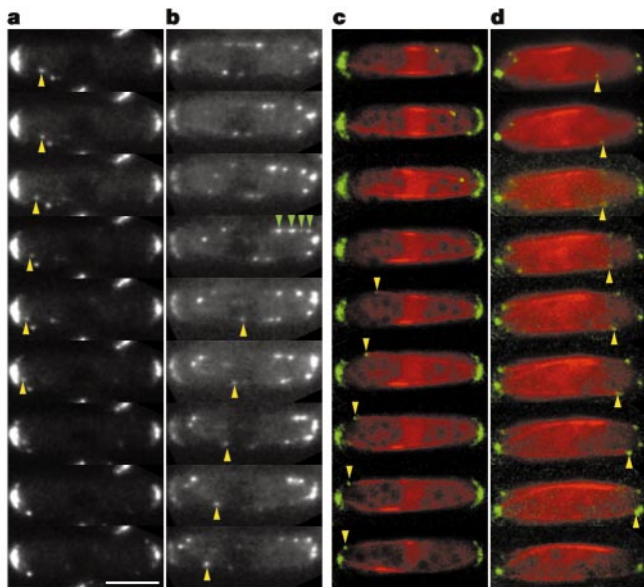


Figure 3 *tea1p* is transported on microtubule ends to cell tips in *mod5Δ* cells. **a, b**, Time lapse series of *tea1p*-GFP in wild-type cells (**a**) and *mod5Δ* cells (**b**). Images in **a** and **b** were taken at 10-s intervals. Yellow arrowheads indicate *tea1p*-GFP particles moving towards cell tips. In **b**, green arrows indicate *tea1p*-GFP spots moving in a stream. **c, d**, Double-labelled time-lapse series of *tea1p*-YFP (green) and CFP-*atb2p* (red) in wild-type cells (**c**) and *mod5Δ* cells (**d**). Images in **c** and **d** were taken at 15-s intervals. QuickTime movie files of this figure are provided in Supplementary Information. Scale bar, 5 μ m.

along the length of a single microtubule bundle rather than being uniquely associated with microtubule ends. However, by imaging tubulin tagged with cyan fluorescent protein (CFP) together with *tea1p* tagged with yellow fluorescent protein (YFP) we found *tea1p*-YFP particles present only at the ends of microtubules in both wild-type and *mod5Δ* cells (Fig. 3c, d; see also Supplementary Information). These results indicate that long microtubule bundles in *mod5Δ* cells might contain some shorter individual microtubules whose ends might not be detectable with current imaging methods. Alternatively, these microtubule ends might exist in wild-type cells as well, but might not normally be occupied by *tea1p* unless the free cytoplasmic pool is increased, as in *mod5Δ* cells. In any case, we do not find evidence for a fundamentally different mode of *tea1p* transport in *mod5Δ* cells.

To determine whether *mod5p* acts locally to retain *tea1p* at cell tips, we investigated its intracellular distribution. Immunostaining of wild-type cells with anti-*mod5p* antibodies showed *mod5p* at cell tips (Fig. 4a; see also Supplementary Information), and this was confirmed in live cells expressing amino-terminally tagged GFP-*mod5p* (Fig. 4b). GFP-*mod5p* localization to cell tips was partly affected by acute disruption of microtubules with MBC (see Supplementary Information) but was unaffected by disruption of actin filaments with latrunculin B (data not shown). The last four amino acids of the predicted *mod5p*-coding sequence are a consensus signal for C-terminal prenylation (CaaX, where C is cysteine, a is an aliphatic amino acid and X is any amino acid); *mod5p* is therefore predicted to be membrane associated¹⁵. We tested the functional significance of this motif both by deleting the last four amino acids of *mod5p* (*mod5ΔCaaXp*) and by mutating the putative prenylated cysteine residue to serine (*mod5C519Sp*); in both cases the normal localization of *mod5p* was disrupted (Fig. 4c; data not shown). *Mod5ΔCaaXp* appeared in a small number of spots throughout the cell, which were typically found near cell tips (Fig. 4c), and in the *mod5ΔCaaX* mutant the localization of *tea1p* was defective (Fig. 4d). These results indicate that the proper

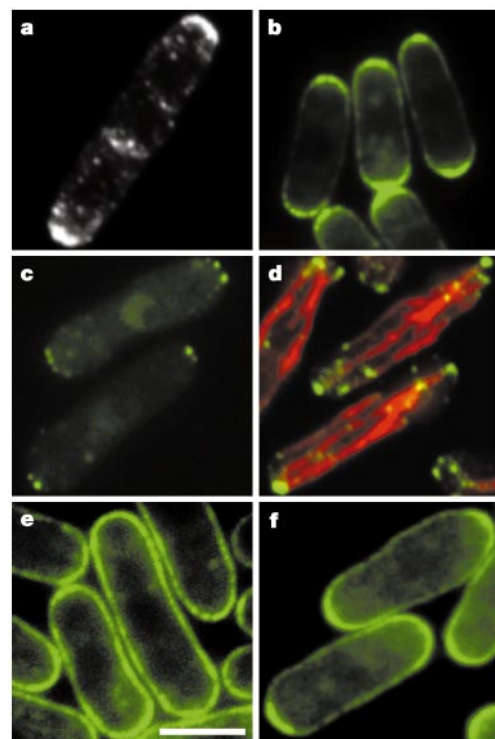


Figure 4 *mod5p* localization to cell tips depends on a functional CaaX sequence and on *tea1p*. **a**, Anti-*mod5p* immunofluorescence in wild-type cells. **b**, Localization of GFP-*mod5p* in wild-type cells (expressed from the *nmt81* promoter, replacing endogenous expression, about 3-fold higher than in the wild type). **c**, Localization of GFP-*mod5ΔCaaXp* (expressed from the *nmt41* promoter, replacing endogenous expression). **d**, Expression of *mod5ΔCaaXp* mimics the phenotype of *mod5Δ*. *mod5ΔCaaX* cells stained with antibodies against *tea1p* (green) and tubulin (red). **e**, Localization of GFP-*mod5p* in *tea1Δ* cells. **f**, Localization of GFP-*mod5p* in *tea3Δ* cells. Scale bar, 5 μ m.

targeting of *mod5p* to the plasma membrane at cell tips might be essential for its function and that a normal microtubule distribution might contribute to its localization.

What *trans*-acting factors might be necessary for localizing *mod5p* at cell tips? We found that in *tea1Δ* cells, GFP-*mod5p* is no longer restricted to the cell tips but instead spreads out across the entire plasma membrane (Fig. 4e). *tea1p* is thought to be important for the cell-tip localization of several other effector proteins involved in polarized growth, including *tip1p*, *tea2p*, *tea3p*, *pom1p* and *bud6p*^{9,10,16–18}. In *tip1Δ* cells, in which short microtubules lead to defects in *tea1p* targeting to cell tips¹⁰, cortical GFP-*mod5p* was similarly mislocalized, but in *tea2-1* mutants, which have a similar microtubule defect, GFP-*mod5p* localization was nearly normal (see Supplementary Information), indicating that the role of *tip1p* in *mod5p* localization might extend beyond the regulation of microtubule dynamics. *tea3Δ* cells showed a slight spreading of GFP-*mod5p* away from cell tips (Fig. 4f; see also Supplementary Information), indicating that *tea3p* might also contribute to restricting *mod5p* to cell tips (see below). In *pom1Δ* and *bud6Δ* strains we observed normal GFP-*mod5p* localization (see Supplementary Information), indicating that they might not be directly involved in the *tea1p*-mediated regulation of *mod5p* localization.

Collectively, these results suggest a model of fission yeast cell polarity regulation involving a positive-feedback loop in which cortically localized *mod5p* at cell tips promotes the anchoring of microtubule-delivered *tea1p*. Correctly anchored *tea1p* then acts reciprocally to prevent *mod5p* from spreading out across the plasma membrane, and this spatial restriction of *mod5p* ultimately leads to

the proper subsequent anchoring of additional tea1p (see Supplementary Information for model). So far we have not been able to demonstrate a biochemical association between mod5p and tea1p, but mod5p is extracted from the cell cortex only under partly denaturing conditions (H.A.S., unpublished observations). In a two-hybrid screen we have identified an interaction between mod5p and tea3p (H.A.S., unpublished observations), which is related in structure to tea1p and interacts with tea1p by two-hybrid analysis, although this has not yet been confirmed biochemically¹⁶. This indicates that mod5p and tea1p might physically interact only indirectly, possibly linked by mutual interactions with tea3p. Recently it has been shown that the deletion of a coiled-coil region at the C terminus of tea1p prevents it from anchoring at the cell cortex⁸; thus, an interaction between mod5p and tea1p might occur through this region of tea1p.

We suggest that the tea1p–mod5p system provides spatially selective anchoring of tea1p at cell tips in the context of dynamic targeting. If tea1p were simply deposited at the cortex after associating with the plus ends of dynamic microtubules, the partly stochastic nature of the orientation of microtubule growth would leave open the possibility of mistargeting, and thus only low-fidelity positioning of tea1p. However, we would argue that this is held in check by the localization interdependence between tea1p and its cortical anchoring factor, mod5p. Interactions between plus-end-associated microtubule-binding proteins and cortical or plasma-membrane proteins, such as the association of EB1 family proteins with adenomatous polyposis coli (APC) protein (or Kar9p in budding yeast)^{19–22}, or of CLIP-170 family proteins with CLIP-associated proteins (CLASPs) and Ras GTPase-activating-protein-like protein (IQGAP1) (refs 23, 24), are thought to be important in a variety of microtubule-based eukaryotic cell behaviours, including mitotic spindle and/or nuclear positioning, and directed cell migration (see ref. 3 for references). It is thought that in most instances the purpose of these interactions is to regulate microtubule dynamics and/or the attachment of microtubules to sub-cellular structures^{2,3,25,26}. In contrast, the primary role of cortical mod5p might be to increase the fidelity of the positioning of microtubule-targeted tea1p in the cell, thus ensuring that the correct identity of the polarized cell tip is continuously asserted. In conjunction with the selective stabilization of microtubule dynamic instability²⁷, spatially selective anchoring of microtubule-targeted proteins might be a more general principle regulating cell polarity. □

Methods

Yeast methods

S. pombe methods were as described²⁸. Mutant cells were identified by starving cells on rich medium plates (YE5S) for 2 days, replica-plating them to fresh plates, and examining their cell shape after 4 h at 32 °C. Deletion of *mod5*⁺ and GFP-tagging of mod5p and tea1p were performed by polymerase chain reaction (PCR)-based gene targeting with the *kanMX* selectable marker²⁹; strains and primers used are listed in Supplementary Information. The C terminus of tea1p was tagged with GFP or YFP. Because of the CaaX box at the C terminus of mod5p, the N terminus of mod5p was tagged with GFP, and GFP–mod5p was expressed under the control of the weak *nmt81* promoter. Before analysis, expression of GFP–mod5p was induced for 2 days at 32 °C in medium lacking thiamine. The *nmt81* promoter drives the expression of GFP–mod5p to a concentration about threefold that of the endogenous protein (see Supplementary Information; additional data not shown). Deletion of the C-terminal CaaX box and mutation of C519S in *mod5* were achieved by PCR-based integration with a *ura4*⁺ selectable marker²⁹, and confirmed by sequencing. Because steady-state expression concentrations of these mutant mod5p proteins from the endogenous *mod5*⁺ promoter are lower than in wild-type cells, expression was restored to wild-type levels by replacing the endogenous promoter with the *nmt41* promoter (data not shown).

In polarity re-establishment experiments, strains were grown to stationary phase in YE5S liquid medium for 3 days at 25 °C before dilution 1:20 into fresh medium at 32 °C, in the presence or absence of 50 µg ml^{−1} MBC⁹. Cells were fixed with formaldehyde and branching was scored by light microscopy.

Antibodies

Anti-tea1p antibodies were raised in sheep against His₆-tea1p (ref. 4). Anti-tea1p antiserum gave only background staining in *tea1Δ* cells (see Supplementary Information),

and was used at a dilution of 1:1,000. TAT1 anti-tubulin hybridoma supernatant³⁰ was a gift from K. Gull, University of Manchester, UK, and used at a dilution of 1:30. Anti-mod5p antibodies were raised in sheep against a GST fusion protein containing amino acids 28–426 of mod5p. The IgG fraction from anti-mod5p antiserum was purified on Protein G–agarose, and used at a dilution of 1:100. Alexa-conjugated fluorescent secondary antibodies (Molecular Probes) were used for immunofluorescence.

Microscopy

For immunostaining with anti-tea1p and anti-tubulin antibodies, cells were harvested by filtration and fixed in methanol at −80 °C for more than 10 min before being processed essentially as described³. Twelve Z-sections at 0.3-µm spacing were captured with a Leica TCS-SP confocal microscope and projected onto a single plane. Because neither mod5p nor GFP–mod5p could be preserved by methanol or aldehyde fixation methods (data not shown), for immunostaining with anti-mod5p antibodies, cells were harvested by filtration and fixed overnight in 10% trichloroacetic acid at 4 °C (ref. 16), then processed as above. Anti-mod5p immunofluorescence images were collected with a Nikon TE300 microscope equipped with Chroma filters, a PiFoc piezofocus device (Physik Instrumente), and a CoolSnapHQ camera (Roper Scientific), controlled by MetaMorph software (Universal Imaging). Seven Z-sections at 0.2-µm spacing were projected onto a single plane. Live-cell imaging of GFP, YFP or CFP was performed on the Nikon TE300 microscope. For time-lapse acquisition of tea1p–GFP, three Z-sections at 0.5-µm spacing were collected with 600-ms exposure time, at 10-s intervals, with appropriate neutral density filters. For time lapse acquisition of cells doubly labelled with tea1p–YFP and CFP–tubulin, two Z-sections at 0.6-µm spacing were collected with exposure times of 1,250 ms for YFP and 1,000 ms for CFP, at 15-s intervals, with neutral density filters. CFP–atb2p was expressed from a plasmid (pRL72 (ref. 18)), a gift from F. Chang, Columbia University, USA) under control of the *nmt1* promoter induced for 20–24 h at 25 °C in 50 nM thiamine.

For measurements of the number of tea1p–GFP particles per cell, particles were counted from Z-stacks containing the entire cell volume; cell tips (terminal 10% of cell length) were not included.

Received 19 December 2002; accepted 7 April 2003; doi:10.1038/nature01672.

- Drubin, D. G. & Nelson, W. J. Origins of cell polarity. *Cell* **84**, 335–344 (1996).
- Schuyler, S. C. & Pellman, D. Microtubule ‘plus-end-tracking proteins’: The end is just the beginning. *Cell* **105**, 421–424 (2001).
- Gundersen, G. Evolutionary conservation of microtubule-capture mechanisms. *Nature Rev. Mol. Cell Biol.* **3**, 296–304 (2002).
- Mata, J. & Nurse, P. Tea1 and the microtubular cytoskeleton are important for generating global spatial order within the fission yeast cell. *Cell* **89**, 939–949 (1997).
- Sawin, K. E. & Nurse, P. Regulation of cell polarity by microtubules in fission yeast. *J. Cell Biol.* **142**, 457–471 (1998).
- Hayles, J. & Nurse, P. A journey into space. *Nature Rev. Mol. Cell Biol.* **2**, 647–656 (2001).
- Chang, F. Establishment of a cellular axis in fission yeast. *Trends Genet.* **17**, 273–278 (2001).
- Behrens, R. & Nurse, P. Roles of fission yeast tea1p in the localization of polarity factors and in organizing the microtubular cytoskeleton. *J. Cell Biol.* **157**, 783–793 (2002).
- Browning, H. et al. Tea2p is a kinesin-like protein required to generate polarized growth in fission yeast. *J. Cell Biol.* **151**, 15–28 (2000).
- Brunner, D. & Nurse, P. CLIP170-like tip1p spatially organizes microtubular dynamics in fission yeast. *Cell* **102**, 695–704 (2000).
- Adachi, Y., Toda, T., Niwa, O. & Yanagida, M. Differential expressions of essential and nonessential alpha-tubulin genes in *Schizosaccharomyces pombe*. *Mol. Cell Biol.* **6**, 2168–2178 (1986).
- Tran, P. T., Marsh, L., Doye, V., Inoue, S. & Chang, F. A mechanism for nuclear positioning in fission yeast based on microtubule pushing. *J. Cell Biol.* **153**, 397–411 (2001).
- Ding, D. Q., Chikashige, Y., Haraguchi, T. & Hiraoka, Y. Oscillatory nuclear movement in fission yeast meiotic prophase is driven by astral microtubules, as revealed by continuous observation of chromosomes and microtubules in living cells. *J. Cell Sci.* **111**, 701–712 (1998).
- Drummond, D. R. & Cross, R. A. Dynamics of interphase microtubules in *Schizosaccharomyces pombe*. *Curr. Biol.* **10**, 766–775 (2000).
- Zhang, F. L. & Casey, P. J. Protein prenylation: Molecular mechanisms and functional consequences. *Annu. Rev. Biochem.* **65**, 241–269 (1996).
- Arellano, M., Niccoli, T. & Nurse, P. Tea3p is a cell end marker activating polarized growth in *Schizosaccharomyces pombe*. *Curr. Biol.* **12**, 751–756 (2002).
- Bahler, J. & Pringle, J. R. Pom1p, a fission yeast protein kinase that provides positional information for both polarized growth and cytokinesis. *Genes Dev.* **12**, 1356–1370 (1998).
- Glynn, J. M., Lustig, R. J., Berlin, A. & Chang, F. Role of bud6p and tea1p in the interaction between actin and microtubules for the establishment of cell polarity in fission yeast. *Curr. Biol.* **11**, 836–845 (2001).
- Bienz, M. The subcellular destinations of APC proteins. *Nature Rev. Mol. Cell Biol.* **3**, 328–338 (2002).
- Miller, R. K., Cheng, S. C. & Rose, M. D. Bim1p/Yeb1p mediates the Kar9p-dependent cortical attachment of cytoplasmic microtubules. *Mol. Biol. Cell* **11**, 2949–2959 (2000).
- Lee, L. et al. Positioning of the mitotic spindle by a cortical-microtubule capture mechanism. *Science* **287**, 2260–2262 (2000).
- Korinek, W. S., Copeland, M. J., Chaudhuri, A. & Chant, J. Molecular linkage underlying microtubule orientation toward cortical sites in yeast. *Science* **287**, 2257–2259 (2000).
- Akhmanova, A. et al. Clasp are CLIP-115 and -170 associating proteins involved in the regional regulation of microtubule dynamics in motile fibroblasts. *Cell* **104**, 923–935 (2001).
- Fukata, M. et al. Rac1 and Cdc42 capture microtubules through IQGAP1 and CLIP-170. *Cell* **109**, 873–885 (2002).
- Tirnauer, J. S., Grego, S., Salmon, E. D. & Mitchison, T. J. EB1-microtubule interactions in *Xenopus* egg extracts: role of EB1 in microtubule stabilization and mechanisms of targeting to microtubules. *Mol. Biol. Cell* **13**, 3614–3626 (2002).

26. Komarova, Y. A., Akhmanova, A. S., Kojima, S., Galjart, N. & Borisy, G. G. Cytoplasmic linker proteins promote microtubule rescue *in vivo*. *J. Cell Biol.* **159**, 589–599 (2002).
27. Kirschner, M. & Mitchison, T. Beyond self-assembly: from microtubules to morphogenesis. *Cell* **45**, 329–342 (1986).
28. Moreno, S., Klar, A. & Nurse, P. Molecular analysis of the fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol.* **194**, 795–823 (1991).
29. Bahler, J. *et al.* Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* **14**, 943–951 (1998).
30. Woods, A. *et al.* Definition of individual components within the cytoskeleton of *Trypanosoma brucei* by a library of monoclonal antibodies. *J. Cell Sci.* **93**, 491–500 (1989).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank members of the Wellcome Trust Centre for Cell Biology for discussions, and R. Allshire, I. Davis, W. Earnshaw and H. Ohkura for critical reading of the manuscript. This work was supported by a Wellcome Trust Senior Research Fellowship to K.E.S. H.A.S. is currently a Caledonian Research Foundation Post-doctoral Fellow.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.E.S. (ken.sawin@ed.ac.uk).

Sphingolipid signalling in *Arabidopsis* guard cells involves heterotrimeric G proteins

Sylvie Coursol^{*†‡}, Liu-Min Fan^{*†}, Hervé Le Stunff^{‡§}, Sarah Spiegel[§], Simon Gilroy^{*} & Sarah M. Assmann^{*}

^{*} Department of Biology, Pennsylvania State University, 208 Mueller Laboratory, University Park, Pennsylvania 16802-5301, USA

[§] Department of Biochemistry, Medical College of Virginia, Virginia Commonwealth University, 1101 East Marshall Street, Sanger Hall, Richmond, Virginia 2398-0614, USA

[†] These authors contributed equally to this work

In animals, the sphingolipid metabolite sphingosine-1-phosphate (S1P) functions as both an intracellular messenger and an extracellular ligand for G-protein-coupled receptors of the S1P receptor family, regulating diverse biological processes ranging from cell proliferation to apoptosis^{1–3}. Recently, it was discovered in plants that S1P is a signalling molecule involved in abscisic acid (ABA) regulation of guard cell turgor⁴. Here we report that the enzyme responsible for S1P production, sphingosine kinase (SphK), is activated by ABA in *Arabidopsis thaliana*, and is involved in both ABA inhibition of stomatal opening and promotion of stomatal closure. Consistent with this observation, inhibition of SphK attenuates ABA regulation of guard cell inward K⁺ channels and slow anion channels, which are involved in the regulation of stomatal pore size. Surprisingly, S1P regulates stomatal apertures and guard cell ion channel activities in wild-type plants, but not in knockout lines of the sole prototypical heterotrimeric G-protein α -subunit gene, *GPA1* (refs 5–8). Our results implicate heterotrimeric G proteins as downstream elements in the S1P signalling pathway that mediates ABA regulation of stomatal function, and suggest that the interplay between S1P and heterotrimeric G proteins represents an evolutionarily conserved signalling mechanism.

In higher plants, stomata form pores on leaf surfaces that regulate

both uptake of CO₂ for photosynthesis and loss of water vapour during transpiration. The plant hormone ABA regulates stomatal pore size by mediating turgor changes in the guard cell pair surrounding the pore. These changes involve alterations in ion channel activities, which represent the final effectors of these processes^{9–11}. Recently, S1P was shown to stimulate stomatal closure and Ca²⁺ mobilization in *Commelina communis* guard cells⁴. In addition, ABA promotion of stomatal closure in this species was partially attenuated by DL-threo-dihydrosphingosine (DL-threo-DHS)⁴, a competitive inhibitor of mammalian SphKs¹². Nevertheless, little is known about the cellular mechanisms that regulate S1P action in plant cells^{13,14}.

We used DL-threo-DHS and N,N-dimethylsphingosine (DMS), another potent inhibitor of mammalian SphKs¹², to investigate whether SphK might be involved in ABA-mediated changes in stomatal apertures in *A. thaliana*. DL-threo-DHS or DMS alone did not affect stomatal apertures (Fig. 1a, d). However, in epidermal peels that had been pre-treated with DL-threo-DHS or DMS, the effect of ABA on both stomatal opening and stomatal closure was significantly attenuated (Fig. 1a, d). Because inhibition of stomatal opening by ABA involves the inhibition of inwardly rectifying K⁺ channels^{10,11}, we next assessed the effect of DMS on inward K⁺ currents in *A. thaliana* guard cell protoplasts (GCPs)^{6,15}. ABA inhibition of inward K⁺ currents was abolished by pre-treating GCPs with DMS (Fig. 1b, c). Because ABA also promotes stomatal closure by activating slow anion channels^{6,11,16}, we next tested whether the effect of ABA on slow anion channels was altered by DMS. ABA activation of slow anion channel activity was partially inhibited by pre-treating GCPs with DMS (Fig. 1e, f). Together, these results suggest that ABA regulation of stomatal apertures and guard cell ion channel activities is transduced, at least in part, through SphK.

SphK activity has yet to be demonstrated in plants, although this enzyme has been characterized in animals¹ and yeast¹⁷. A related kinase activity able to phosphorylate D-erythro-dihydrosphingosine (D-erythro-DHS), which is structurally similar to D-erythro-sphingosine (Sph) except for the absence of the *trans*-4,5 double bond, has been reported in *Zea mays*¹⁸. A recombinant long-chain sphingoid base kinase (AtLCBK1) from *A. thaliana* was also shown to phosphorylate D-erythro-DHS¹⁹. Here, we show by direct assays of sphingoid base phosphorylation²⁰ that SphK activity is present in lysates prepared from leaves, mesophyll cell protoplasts (MCPs), or GCPs of *A. thaliana* (Fig. 2a). We found that Sph, a low abundance, naturally occurring sphingoid base isomer in *A. thaliana* leaves²¹, was the best substrate among various exogenous sphingoid bases tested (Fig. 2b). Of the known inhibitors of mammalian SphKs, DL-threo-DHS was a substrate for *A. thaliana* SphK, whereas DMS was not (Fig. 2b). Furthermore, we found that DMS inhibited the activity of *A. thaliana* SphK in a concentration-dependent manner (Fig. 2c). Although a higher concentration of DMS was required for inhibition *in vitro*, such differences between the *in vitro* and *in vivo* concentrations are similar to those reported for other inhibitors used. For example, the phospholipase C (PLC) inhibitor U-73122 required an 80-fold higher concentration than was effective *in vivo* to inhibit guard cell PLC activity *in vitro*²². These results, together with those from Fig. 1, support the involvement of SphK in the regulation of guard cell responses to ABA.

We next tested the ability of ABA to modulate S1P levels *in vivo* by measuring the formation of [³H]S1P from [³H]Sph. Treatment of GCPs and MCPs with ABA resulted in a rapid and transient increase in [³H]S1P formation, reaching a maximum after 2 min of ABA treatment (Fig. 2d, e). As expected, pre-treatment of MCPs with DMS blocked the ABA-induced increases in [³H]S1P formation (data not shown). We next assayed SphK activity *in vitro* from protein extracts of MCPs that had been treated with ABA for various times *in vivo*. We found that stimulation of MCPs with ABA rapidly and transiently increased SphK activity (Fig. 2f), with a time course

[‡] Present address: Station de Génétique Végétale, UMR 320 INRA/8120 CNRS/Université Paris-Sud/INA-PG, Ferme du Moulon, 91190 Gif-sur-Yvette, France (S.C.); and Laboratoire d'Activation Cellulaire et Transduction des Signaux, Institut de Biochimie et de Biophysique Moléculaire et Cellulaire, UMR 8619 CNRS, Université Paris-Sud, 91405 Orsay Cedex, France (H.L.S.).

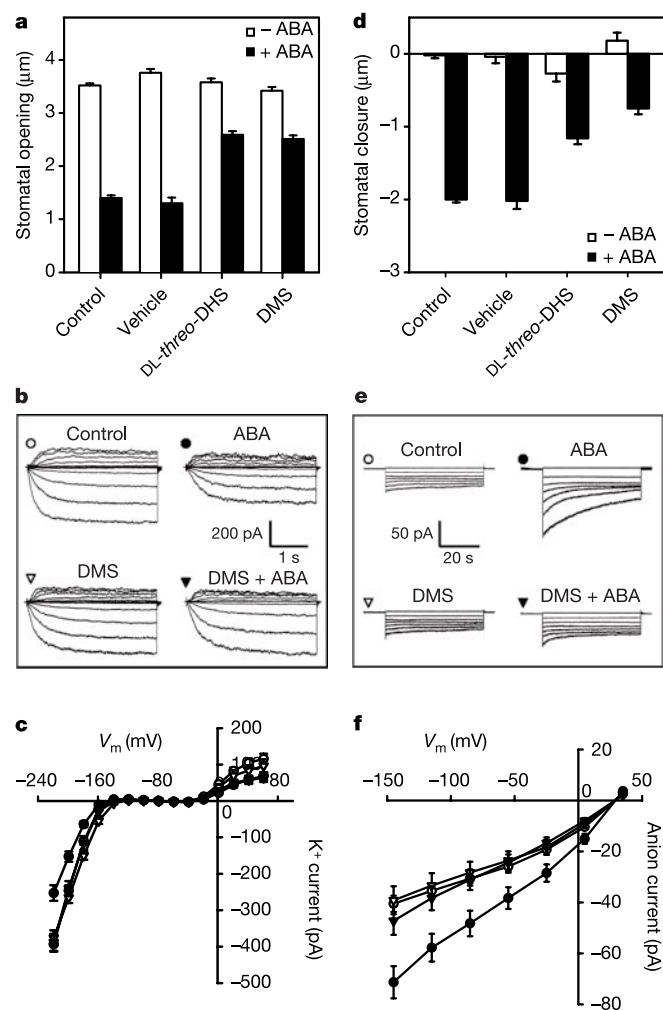


Figure 1 Effect of SphK inhibitors on ABA regulation of stomatal apertures and guard cell ion channel activities. **a**, Apertures after a 2.5 h pre-treatment in the dark in 20 μ M DL-threo-DHS or 5 μ M DMS (vehicle, dimethylsulphoxide; final concentration, $\leq 0.08\%$), and a further 2 h light treatment with or without 20 μ M ABA. **b**, Representative whole-cell recordings of K^+ currents from wild-type GCPs pre-treated with 5 μ M DMS for 30 min, and subsequently treated for 1.5 h with or without 50 μ M ABA; 50 μ M ABA was also included in bath solution. **c**, Current-voltage curves of time-activated whole-cell K^+ currents in wild-type GCPs; $n = 31, 9, 21$ and 16 cells for control, DMS, ABA and DMS + ABA treatments, respectively. **d**, Apertures after a 2.5 h light pre-treatment in 20 μ M DL-threo-DHS or 5 μ M DMS, and a further 3 h light treatment with or without 20 μ M ABA. **e**, Representative whole-cell recordings of slow anion currents from wild-type GCPs treated as in **b**, except that 50 μ M ABA was also included in the pipette solution¹⁶ in ABA and DMS + ABA treatments. **f**, Current-voltage curves of steady-state whole-cell anion currents in wild-type GCPs; $n = 31, 13, 24$ and 14 cells for control, DMS, ABA and DMS + ABA treatments, respectively.

similar to that of ABA-increased [3 H]S1P formation *in vivo* (Fig. 2d, e). Collectively, these results provide evidence that ABA increases the activity of SphK, leading to the production of S1P in plants. To investigate further the nature of the interaction between ABA and SphK, we assayed SphK stimulation by ABA in cell-free systems derived from leaves, MCPs or GCPs. We found that ABA increased the phosphorylation of both exogenously supplied Sph and an endogenous long-chain sphingoid base (see Supplementary Information). These results, together with those of Fig. 2, suggest that the ABA receptor, signal transduction system^{9–11} and ABA-regulated SphK activity might exist in a

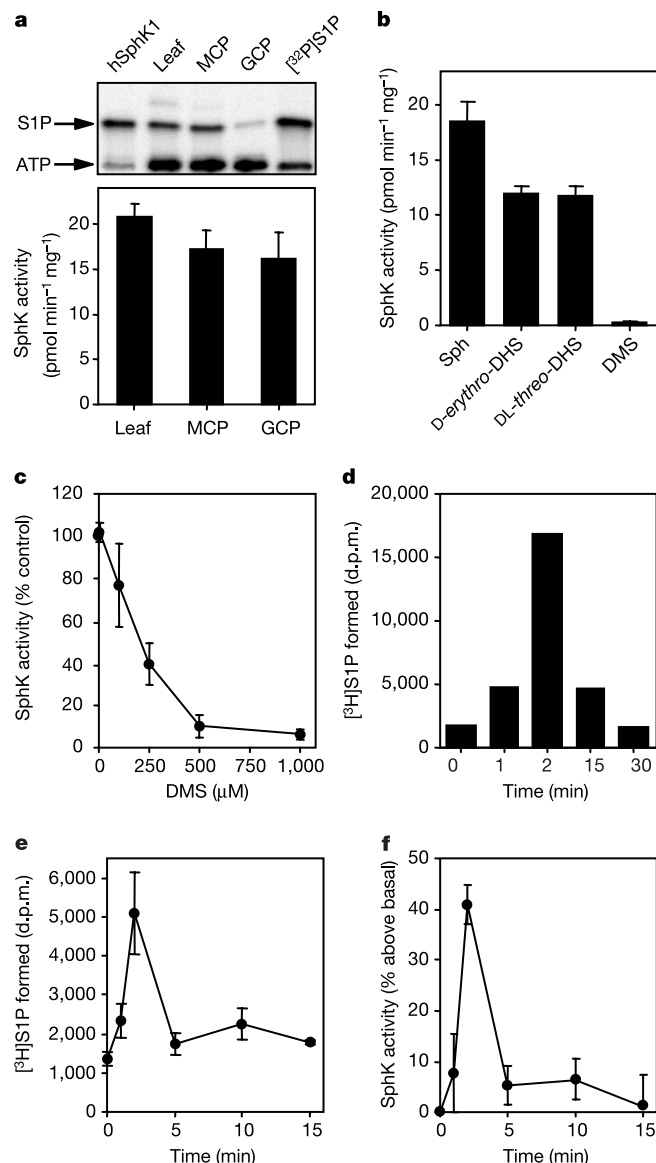


Figure 2 Effect of ABA on SphK activity. **a**, Thin layer chromatography (TLC) analysis of [32 P]S1P formed by lysates from leaves (25 μ g), MCPs (25 μ g), GCPs (3 μ g) and recombinant human SphK1 (hSphK1) with 50 μ M Sph as substrate (top panel). Without Sph added in the *in vitro* assay, no S1P was formed. The bottom panel shows the specific activity of SphK. **b**, Phosphorylation of various sphingoid bases (50 μ M) by leaf lysate. **c**, Effect of various concentrations of DMS on SphK activity in leaf lysate with 5 μ M Sph as substrate. **d**, Stimulation of [3 H]S1P formation *in vivo* by 50 μ M ABA in GCPs. The figure shows data representative of one of three separate experiments. **e**, Stimulation of [3 H]S1P formation *in vivo* by 50 μ M ABA in MCPs. Values are the mean $\pm$ s.e.m. from four independent experiments and data were compared using the Student's *t*-test at the 95% confidence level. **f**, SphK activity assayed *in vitro* from protein extracts of MCPs that had been treated with 50 μ M ABA for various time *in vivo*.

tightly associated complex that survives the extraction process required for the *in vitro* assay.

In animal cells, an extracellular action of S1P by way of heterotrimeric G-protein-coupled receptors (GPCRs) for S1P is well established^{1–3}. Recently, we used two independent *A. thaliana* transferred DNA (T-DNA) knockout lines, *gpa1-1* and *gpa1-2* (ref. 5), to provide evidence that GPA1 is involved in ABA signalling in *A. thaliana* guard cells⁶. Therefore, we examined whether S1P signals in guard cells can be transduced through GPA1, which in

turn regulates K^+ and slow anion channel activities and stomatal apertures⁶. We found that basal levels of SphK activity were similar in leaf and GCP lysates of wild-type and *gpa1* null mutant plants (data not shown). In wild-type plants, S1P inhibited the opening of closed stomata to a similar extent as ABA (Fig. 3a). In addition, S1P promoted the closure of opened stomata (Fig. 3d), in agreement with previous observations in *C. communis* guard cells⁴. By contrast, in *gpa1* null mutants, S1P neither inhibited stomatal opening nor promoted stomatal closure (Fig. 3a, d), although stomatal closure (Fig. 3d) in these genotypes occurs normally in response to ABA⁶. These results, together with those of Fig. 1, suggest that GPA1 functions downstream of SphK and S1P in ABA signal transduction pathways that lead to changes in stomatal apertures in *A. thaliana*. Studies using animal cells have shown that D-erythro-dihydro-

sphingosine-1-phosphate (dihydro-S1P), which is structurally similar to S1P except for the absence of the *trans*-4,5 double bond, is a ligand for S1P receptors and activates the same downstream responses as S1P²³. However, in contrast to S1P, dihydro-S1P neither promoted stomatal closure (Fig. 3d), as previously reported for *C. communis* guard cells⁴, nor inhibited stomatal opening (Fig. 3a) in wild-type and *gpa1* null mutant plants.

We next tested the ability of S1P to modulate guard cell ion channel activities in wild-type and *gpa1* null mutant plants. In mammalian cells, S1P activates muscarinic receptor-coupled inwardly rectifying K^+ channels, probably by way of the S1P₃ GPCR²⁴, and Ca^{2+} -dependent K^+ channels²⁵. By contrast, our data show that S1P inhibits the inward K^+ channels in GCPs from wild-type plants. Importantly, this inhibitory effect is dependent on heterotrimeric G proteins, as evident from the lack of S1P inhibition of inward K^+ channels in *gpa1* null mutants (Fig. 3b, c). We next examined the effect of S1P on slow anion channels, which are activated by ABA, in part through heterotrimeric G proteins⁶. We found that S1P stimulated slow anion currents in guard cells of wild-type plants, but not in *gpa1* null mutant lines (Fig. 3e, f). This result is consistent with the absence of S1P stimulation of stomatal closure in these mutant lines (Fig. 3d). Together, these results highlight the central role of GPA1 in mediating the diametrically

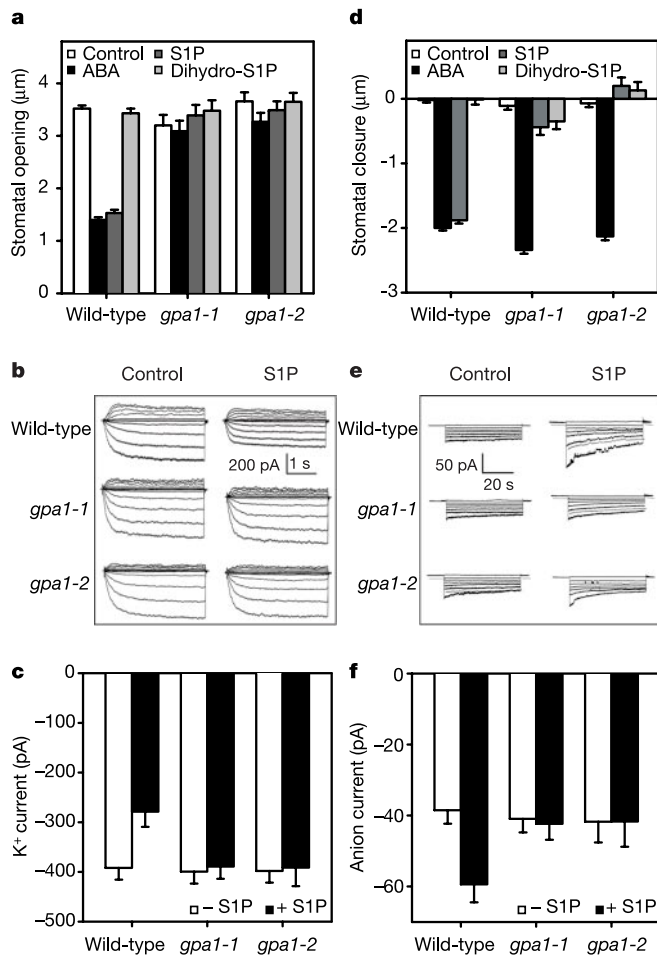


Figure 3 Effects of S1P on stomatal apertures and guard cell ion channel activities in wild-type and G-protein α -subunit (G α) null mutant *gpa1-1* and *gpa1-2* plants. **a**, Apertures after a 2.5 h pre-treatment in the dark, and a further 2 h light treatment with or without 20 μ M ABA, 10 μ M S1P or 10 μ M dihydro-S1P. **b**, Representative whole-cell recordings of K^+ currents from wild-type, *gpa1-1* and *gpa1-2* GCPs. Control recordings were obtained 10 min after achieving the whole-cell configuration. GCPs were subsequently challenged with 10 μ M S1P, and recordings were acquired again after 20 min. **c**, Average time-activated whole-cell K^+ currents at -219 mV; $n = 18$, 10 and 7 cells for wild-type, *gpa1-1* and *gpa1-2*, respectively. **d**, Apertures after a 2.5 h light pre-treatment, and a further 3 h light treatment with or without 20 μ M ABA, 10 μ M S1P or 10 μ M dihydro-S1P. **e**, Representative whole-cell recordings of slow anion currents from wild-type, *gpa1-1* and *gpa1-2* GCPs. GCPs were pre-treated with 10 μ M S1P for 30 min; 10 μ M S1P was also added to bath solution. **f**, Average steady-state whole-cell anion currents at -145 mV; $n = 25$, 24 and 23 cells, and $n = 29$, 29 and 23 cells for wild-type, *gpa1-1* and *gpa1-2* treated with or without S1P, respectively.

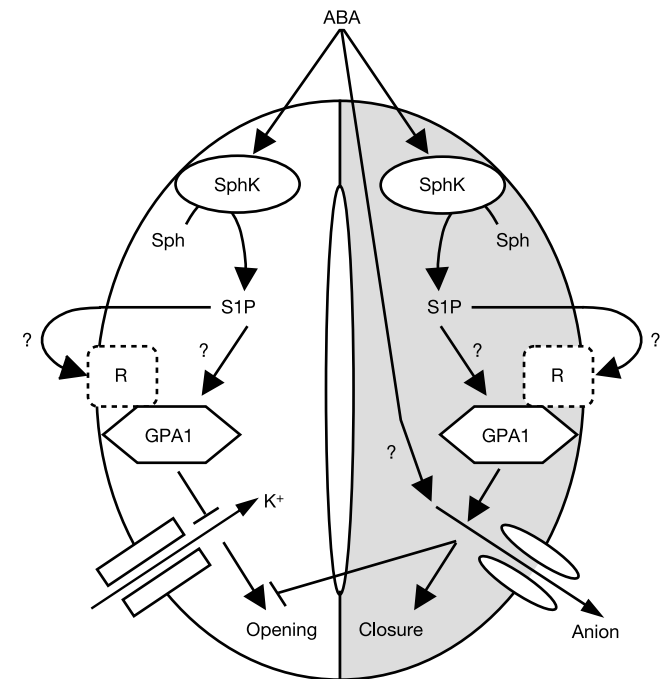


Figure 4 Model of ABA activation of S1P signalling in *A. thaliana* guard cells. The model integrates only the results presented here and previous data on *gpa1* knockout plants⁹. The other signalling elements known to operate in these cells⁹⁻¹¹ are omitted from this S1P-related model for the sake of clarity. S1P formed by ABA activation of SphK may function as a ligand for an unknown receptor (R), or may interact directly with GPA1 to inhibit and activate plasma membrane inwardly rectifying K^+ channels and slow anion channels, respectively, resulting in inhibition of stomatal opening and promotion of stomatal closure. Major points that remain to be clarified by further experiments are indicated by question marks. Because ABA activates stomatal closure and slow anion channels in *gpa1* knockout lines but S1P does not, the solid line is added to represent an S1P-independent mechanism of ABA action on stomatal closure, the precise nature of which requires further investigation. Possible interactions between the illustrated components and the diverse array of other intermediates known to be involved in guard cell ABA signalling⁹⁻¹¹ remain to be elucidated.

opposing effects of S1P on guard cell inward K^+ channels and slow anion channels.

In this study, we have provided a biochemical basis for how S1P is generated in guard cells in response to ABA, and have elucidated the underlying mechanisms whereby the S1P signal is transduced within the cellular context to mediate changes in guard cell turgor (Fig. 4). Future work may allow the GPA1-dependent pathways of S1P action (Fig. 4) to be linked with other known ABA signalling components, such as reactive oxygen species, cytosolic pH and levels of cytosolic Ca^{2+} (refs 9–11). Furthermore, it will be important to elucidate the molecular mechanisms by which S1P affects GPA1. In animal cells, S1P and dihydro-S1P act as extracellular ligands for the S1P receptor family^{1,23}. In guard cells, S1P signal transduction by way of a GPA1-coupled S1P-like receptor is less likely, because dihydro-S1P did not affect stomatal responses (Fig. 3a, d). Moreover, the putative GPCR of *A. thaliana*, GCR1^{7,8,26,27}, bears no sequence homology to any of the conserved S1P receptors. Therefore, the S1P signal in guard cells may be transduced either through a structurally unique GPCR, by a direct interaction of S1P with GPA1, or through unidentified proteins that function to input signals to heterotrimeric G proteins independently of GPCRs. This latter possibility is lent credence by recent evidence in mammalian cells that AGS (activator of G-protein signalling) proteins can activate heterotrimeric-G-protein signalling pathways independently of GPCRs^{7,8,28}, and might help to explain the enigmatic intracellular action of S1P in mammalian cell growth, given that proteins serving as intracellular targets of S1P have not yet been identified^{1,23}. Future work is likely to reveal novel mechanisms of S1P signalling through heterotrimeric G proteins in diverse eukaryotes. □

Methods

Preparation of GCPs and MCPs

GCPs were isolated as described^{6,15}. MCPs were isolated by modification of a method described previously¹⁵ (see Supplementary Methods).

Stomatal bioassay and electrophysiology

Stomatal aperture bioassays and whole-cell current recordings were conducted as described⁶, with minor modifications (see Supplementary Methods).

Preparation of lysates from leaves, MCPs and GCPs

All procedures were conducted at 4 °C. Rosette leaves of 4- to 5-week-old plants^{6,15} were homogenized with SphK buffer²⁰ containing 0.6% (w/v) insoluble polyvinylpyrrolidone (Sigma) and sand (–50 +70 Mesh, Sigma). The homogenate was then centrifuged at 10,000g for 10 min. The supernatant was recovered, frozen with liquid N_2 , and stored at –80 °C. Lysates from GCPs and MCPs were prepared by collecting and resuspending protoplasts in 400 μ l of SphK buffer²⁰. Protoplasts were homogenized and centrifuged at 10,000g for 10 min. The supernatants were then frozen with liquid N_2 , and stored at –80 °C.

SphK assays

Aliquots of the MCP suspension (100 μ l, approximately 7.5×10^5 protoplasts) were incubated at room temperature for 5 min, and treated with or without 50 μ M ABA prepared as described¹⁵. After the appropriate incubation time, each sample was homogenized in a glass/Teflon homogenizer with 300 μ l of SphK buffer²⁰, and centrifuged at 10,000g for 10 min. The supernatants were then frozen with liquid N_2 , and stored at –80 °C. SphK activity was assayed by providing [γ -³²P]ATP and Sph (see Supplementary Methods) as substrates to the lysates, and measuring radiolabelled S1P formation²⁰. Values are the mean $\pm$ s.e.m. from three to five independent experiments, and data were compared using the Student's *t*-test at the 95% confidence level.

Measurement of [³H]S1P formation

Intracellular levels of [³H]S1P were measured by modification of a method described previously²⁹. Aliquots of MCP and GCP suspensions (200 μ l, approximately 5×10^5 protoplasts) were incubated at 25 °C for 10 min, before the addition of 30 nM [³H]Sph (20 Ci mmol^{–1}) and 0.2 mg ml^{–1} fatty-acid-free BSA for 1 min (approximately 300,000 d.p.m. per assay). Reactions were started by the addition of 50 μ M ABA prepared as described¹⁵. After the appropriate incubation time, reactions were terminated by the addition of 500 μ l of chloroform/methanol/concentrated HCl (100:200:1, v/v/v) and 40 μ l of 2 M HCl. After vortexing, 250 μ l of chloroform and 250 μ l of 2 M KCl were added. The samples were mixed thoroughly, followed by centrifugation at 2,000g for 5 min. The resultant organic phases were collected, concentrated under a stream of N_2 , and spotted onto Silica gel 60 thin-layer chromatography (TLC) plates. Samples were run in a 1-butanol/

acetic acid/water (3:1:1, v/v/v) solvent system, and Sph and S1P identified with ninhydrin. After scraping off the spots corresponding to [³H]S1P, radioactivity was measured by liquid scintillation counting (Optiphase Hisafe 3, Wallac).

Received 5 February; accepted 7 April 2003; doi:10.1038/nature01643.

1. Spiegel, S. & Milstien, S. Sphingosine-1-phosphate, a key cell signaling molecule. *J. Biol. Chem.* **277**, 25851–25854 (2002).
2. Pyne, S. & Pyne, N. J. Sphingosine 1-phosphate signalling and termination at lipid phosphate receptors. *Biochim. Biophys. Acta* **1582**, 121–131 (2002).
3. Hla, T., Lee, M.-J., Ancellin, N., Paik, J. H. & Kluck, M. J. Lysophospholipids—receptor revelations. *Science* **294**, 1875–1878 (2001).
4. Ng, C. K.-Y., Carr, K., McAnish, M. R., Powell, B. & Hetherington, A. M. Drought-induced guard cell signal transduction involves sphingosine-1-phosphate. *Nature* **410**, 596–599 (2001).
5. Ullah, H. *et al.* Modulation of cell proliferation by heterotrimeric G protein in *Arabidopsis*. *Science* **292**, 2066–2069 (2001).
6. Wang, X.-Q., Ullah, H., Jones, A. M. & Assmann, S. M. G protein regulation of ion channels and abscisic acid signaling in *Arabidopsis* guard cells. *Science* **292**, 2070–2072 (2001).
7. Jones, A. M. G-protein-coupled signaling in *Arabidopsis*. *Curr. Opin. Plant Biol.* **5**, 402–407 (2002).
8. Assmann, S. M. Heterotrimeric and unconventional GTP binding proteins in plant cell signaling. *Plant Cell* **14**, S355–S373 (2002).
9. Assmann, S. M. & Wang, X.-Q. From milliseconds to millions of years: guard cells and environmental responses. *Curr. Opin. Plant Biol.* **4**, 421–428 (2001).
10. Blatt, M. R. Cellular signaling and volume control in stomatal movements in plants. *Annu. Rev. Cell Dev. Biol.* **16**, 221–241 (2000).
11. Schroeder, J. I., Kwak, J. M. & Allen, G. J. Guard cell abscisic acid signalling and engineering drought hardiness in plants. *Nature* **410**, 327–330 (2001).
12. Kohama, T. *et al.* Molecular cloning and functional characterization of murine sphingosine kinase. *J. Biol. Chem.* **273**, 23722–23728 (1998).
13. Brownlee, C. Intracellular signalling: sphingosine-1-phosphate branches out. *Curr. Biol.* **11**, R535–R538 (2001).
14. Ng, C. K.-Y. & Hetherington, A. M. Sphingolipid-mediated signalling in plants. *Ann. Bot. (Lond.)* **88**, 957–965 (2001).
15. Pandey, S., Wang, X.-Q., Coursol, S. & Assmann, S. M. Preparation and applications of *Arabidopsis thaliana* guard cell protoplasts. *New Phytol.* **153**, 517–526 (2002).
16. Pei, Z.-M., Kuchitsu, K., Ward, J. M., Schwarz, M. & Schroeder, J. I. Differential abscisic acid regulation of guard cell slow anion channels in *Arabidopsis* wild-type and *abi1* and *abi2* mutants. *Plant Cell* **9**, 409–423 (1997).
17. Dickson, R. C. & Lester, R. L. Sphingolipid functions in *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta* **1583**, 13–25 (2002).
18. Crowther, G. J. & Lynch, D. V. Characterization of sphinganine kinase activity in corn shoot microsomes. *Arch. Biochem. Biophys.* **337**, 284–290 (1997).
19. Nishiura, H., Tamura, K., Morimoto, Y. & Imai, H. Characterization of sphingolipid long-chain base kinase in *Arabidopsis thaliana*. *Biochem. Soc. Trans.* **28**, 747–748 (2000).
20. Olivera, A., Kohama, T., Tu, Z., Milstien, S. & Spiegel, S. Purification and characterization of rat kidney sphingosine kinase. *J. Biol. Chem.* **273**, 12576–12583 (1998).
21. Sperling, P., Zähringer, U. & Heinz, E. A sphingolipid desaturase from higher plants. Identification of a new cytochrome *b₅* fusion protein. *J. Biol. Chem.* **273**, 28590–28596 (1998).
22. Staxén, I. *et al.* Abscissic acid induces oscillations in guard-cell cytosolic free calcium that involve phosphoinositide-specific phospholipase C. *Proc. Natl Acad. Sci. USA* **96**, 1779–1784 (1999).
23. van Brocklyn, J. R. *et al.* Dual actions of sphingosine-1-phosphate: extracellular through the G_i-coupled receptor Edg-1 and intracellular to regulate proliferation and survival. *J. Cell Biol.* **142**, 229–240 (1998).
24. Himmel, H. M. *et al.* Evidence for EDG-3 receptor-mediated activation of *I(K_{ACh})* by sphingosine-1-phosphate in human atrial cardiomyocytes. *Mol. Pharmacol.* **58**, 449–454 (2000).
25. Schilling, T. *et al.* Lysophospholipids induce membrane hyperpolarization in microglia by activation of IKCA1 Ca^{2+} -dependent K^+ channels. *Neuroscience* **109**, 827–835 (2002).
26. Joseffson, L. G. & Rask, L. Cloning of a putative G-protein-coupled receptor from *Arabidopsis thaliana*. *Eur. J. Biochem.* **249**, 415–420 (1997).
27. Colucci, G., Apone, F., Alyshemmi, N., Chalmers, D. & Chrispeels, M. J. *GCR1*, the putative *Arabidopsis* G protein-coupled receptor gene is cell cycle-regulated, and its overexpression abolishes seed dormancy and shortens time to flowering. *Proc. Natl Acad. Sci. USA* **99**, 4736–4741 (2002).
28. Cismowski, M. J., Takesono, A., Bernard, M. L., Duzic, E. & Lanier, S. M. Receptor-independent activators of heterotrimeric G-proteins. *Life Sci.* **68**, 2301–2308 (2001).
29. Meyer zu Heringdorf, D. *et al.* Sphingosine kinase-mediated Ca^{2+} signalling by G-protein-coupled receptors. *EMBO J.* **17**, 2830–2837 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. P. Hobson for the recombinant purified hSphK1, X.-Q. Wang, A. Olivera and J.-N. Pierre for technical advice, J. Coursol for statistical analyses, and T. Jacob and C. K.-Y. Ng for critically reading the manuscript. This work was supported by grants from the United States Department of Agriculture (USDA) and the National Science Foundation to S.M.A., from the USDA to S.G., and from the National Institutes of Health to S.S.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.M.A. (sma3@psu.edu).

Histone H3 phosphorylation by IKK- α is critical for cytokine-induced gene expression

Yumi Yamamoto, Udit N. Verma, Shashi Prajapati, Youn-Tae Kwak & Richard B. Gaynor

Division of Hematology-Oncology, Department of Medicine, Harold Simmons Cancer Center, University of Texas Southwestern Medical Center, Dallas, Texas 75390-8594, USA

Cytokine-induced activation of the I κ B kinases (IKK) IKK- α and IKK- β is a key step involved in the activation of the NF- κ B pathway¹⁻⁴. Gene-disruption studies of the murine IKK genes have shown that IKK- β , but not IKK- α , is critical for cytokine-induced I κ B degradation⁵⁻⁷. Nevertheless, mouse embryo fibroblasts deficient in IKK- α are defective in the induction of NF- κ B-dependent transcription⁷⁻⁹. These observations raised the question of whether IKK- α might regulate a previously undescribed step to activate the NF- κ B pathway that is independent of its previously described cytoplasmic role in the phosphorylation of I κ B α . Here we show that IKK- α functions in the nucleus to activate the expression of NF- κ B-responsive genes after stimulation with cytokines. IKK- α interacts with CREB-binding protein and in conjunction with Rel A is recruited to NF- κ B-responsive promoters and mediates the cytokine-induced phosphorylation and subsequent acetylation of specific residues in histone H3. These results define a new nuclear role of IKK- α in modifying histone function that is critical for the activation of NF- κ B-directed gene expression.

First, we characterized cytokine-induced activation of the NF- κ B pathway in parental, IKK- α ^{-/-} and IKK- β ^{-/-} mouse embryo fibroblasts (MEFs). Tumour-necrosis factor- α (TNF- α) induced the rapid degradation of I κ B α in both parental MEF cells and in IKK- α ^{-/-} cells but not in IKK- β ^{-/-} cells (Fig. 1a). However, defects in the expression of an NF- κ B reporter were noted after treatment with TNF- α in both IKK- α ^{-/-} and IKK- β ^{-/-} cells without changes in the expression of a Rous sarcoma virus- β -galactosidase reporter (Fig. 1b). Furthermore, quantitative real-time polymerase chain reaction (PCR) analysis confirmed that TNF- α increased the transcription of the NF- κ B-regulated I κ B α gene in both MEF and IKK- α ^{+/-} cells, in which Myc-tagged IKK- α was stably expressed in IKK- α ^{-/-} cells but not in IKK- α ^{-/-} and IKK- β ^{-/-} cells (Fig. 1c). These results indicated that IKK- α is critical for the rapid cytokine-mediated induction of NF- κ B-responsive genes by a mechanism distinct from that of IKK- β -mediated increases in I κ B degradation.

It has recently been noted that IKK- α can shuttle between cytoplasm and nucleus¹⁰. To address whether the intracellular distribution of IKK- α and IKK- β was different, these endogenous proteins were immunostained in HeLa cells (Fig. 1d) in addition to both parental and IKK-deficient mouse embryo fibroblasts (Fig. 1e). IKK- α had a different distribution from that of IKK- β : IKK- α was localized in both the cytoplasm and the nucleus, whereas IKK- β was localized predominantly in the cytoplasm (Fig. 1d, e). These results indicate that IKK- α might have an additional nuclear role in activating the NF- κ B pathway that is distinct from its previously described role in phosphorylating the I κ B proteins.

To investigate whether IKK- α might function in the nucleus to regulate the cytokine-induced expression of the NF- κ B-responsive I κ B α and interleukin-8 (IL-8) genes, chromatin immunoprecipitation (ChIP) assays were used. HeLa cells were treated with TNF- α and ChIP assays were performed at various times after stimulation. In response to stimulation with TNF- α , the recruitment of IKK- α ,

p65 and CREB binding protein (CBP) to the I κ B α and IL-8 promoters was detected 15 minutes after treatment and was present for at least 120 minutes (Fig. 2a, left and middle panels). In contrast, there was no association of either IKK- β or IKK- γ /NEMO with these promoters unless several additional PCR cycles were used (Fig. 2a, left and middle panels). There was no cytokine-induced association of any of these factors with the β -actin promoter (Fig. 2a, right panel). Quantitative real-time PCR analysis was used to determine the percentage of I κ B α and IL-8 promoter input DNA that was bound by IKK- α , p65 and CBP (Fig. 2b). Finally, the kinetics of binding of these factors to the I κ B α and IL-8 genes was correlated with their increased transcription (Fig. 2c).

Given the fact that interactions between CBP and p65 have been shown to be important for NF- κ B activation¹¹⁻¹⁴, it was possible that IKK- α might form a complex with these proteins on NF- κ B-regulated promoters to increase transcription. IKK- α might then activate NF- κ B-regulated transcription by the phosphorylation of p65 (refs 15, 16) or potentially other nuclear targets. First, the interactions of epitope-tagged proteins including CBP and either IKK- α , IKK- β or p65 were analysed in extracts prepared from cells co-transfected with expression vectors encoding these proteins. Immunoprecipitation and subsequent western blot analysis

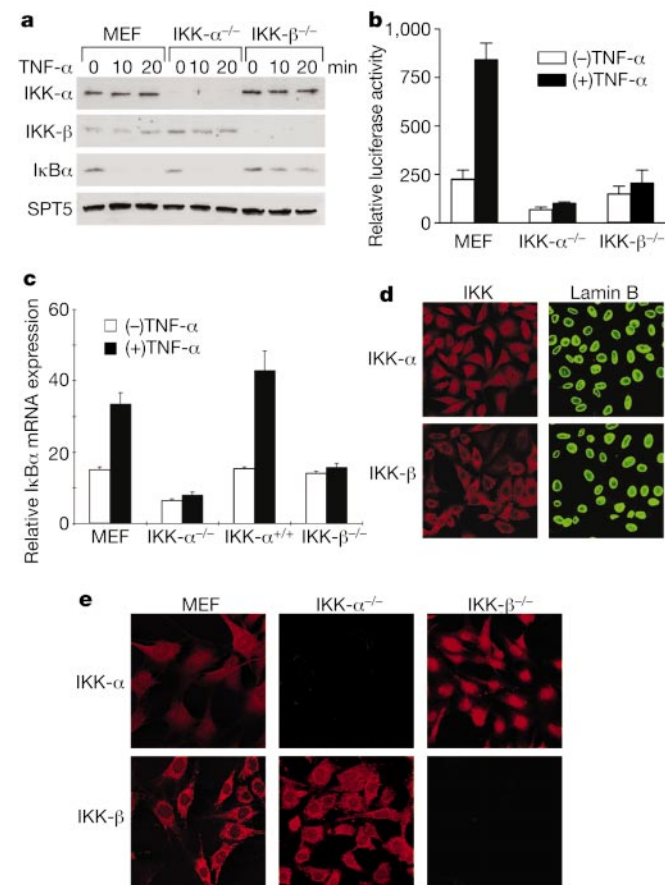


Figure 1 Both IKK- α ^{-/-} and IKK- β ^{-/-} cells are defective in TNF- α -induced NF- κ B activation. **a**, Cells were treated with TNF- α for the specified durations, and western blot analysis of these extracts was performed with the indicated antibodies. **b**, TNF- α -induced NF- κ B activation was determined by luciferase assays after transfection of an NF- κ B-dependent reporter construct and normalized by a Rous sarcoma virus- β -galactosidase reporter construct. **c**, The indicated cell lines were either untreated or treated with TNF- α for 15 min and the change in I κ B α mRNA expression was determined by quantitative real-time PCR and normalized by measuring the levels of 18S RNA. **d**, Immunofluorescence of HeLa cells and **e**, of MEF, IKK- α ^{-/-} and IKK- β ^{-/-} cells was performed with the indicated IKK- α -specific and IKK- β -specific antibodies.

demonstrated that there were strong interactions between IKK- α and CBP as well as between p65 and CBP (Fig. 3a). In addition, the mammalian two-hybrid assay demonstrated that IKK- α strongly interacted with the amino-terminal transactivation domain of CBP, which is also the binding site for p65 (Fig. 3b)¹⁷. However, no direct interaction was detected between IKK- α and p65 in this analysis (data not shown). Consistent with the two-hybrid data is the observation that IKK- α bound to a glutathione *S*-transferase (GST)–CBP fusion protein containing the N terminus of CBP (Fig. 3c). Similar data demonstrating the interaction of IKK- α and p65, but not IKK- β , with CBP were also seen with HeLa extracts containing either these transiently expressed proteins (Fig. 3d), or endogenous proteins (Fig. 3e).

It was important to determine whether p65 was required for the association of IKK- α with the I κ B α promoter. For these studies we used infection with adenoviruses encoding either the I κ B super-repressor^{18,19}, which expresses an I κ B protein that is not degraded after stimulation by cytokine and prevents both p65 nuclear translocation and DNA binding, or β -galactosidase as a control (Fig. 4a). ChIP assays demonstrated that the I κ B super-repressor prevented the association of p65, but not IKK- α or CBP, with the I κ B promoter (Fig. 4a). Although these results suggested that p65

was not crucial for IKK- α association, we found defects in the association of both IKK- α and CBP, in addition to other factors that bound to the I κ B α and IL-8 promoter, in ChIP assays performed in TNF- α -treated p65^{-/-} cells (data not shown). It remains to be determined whether the defect in IKK- α and CBP promoter recruitment seen in p65^{-/-} cells was due only to the failure of p65 to bind to NF- κ B-regulated promoters or to other effects.

Next we addressed whether intact IKK- α kinase activity was required for association with the I κ B α promoter. Both the epitope-tagged wild-type and mutant IKK- α proteins were recruited to the I κ B α promoter after treatment with TNF- α (Fig. 4b). The ability of these IKK- α proteins to alter CBP-mediated gene expression directly was next addressed. Transfection of wild-type IKK- α , but not the kinase-dead IKK- α (K/M) mutant or IKK- β , dose-dependently enhanced GAL4–CBP-mediated transcriptional activity (Fig. 4c). These results suggested that although both wild-type and mutant IKK- α proteins could associate with the I κ B α promoter, IKK- α kinase activity was important for modulating CBP-dependent transcription.

These observations raised the question of the mechanism by which IKK- α could modulate CBP-dependent transcriptional activation. The histone acetyltransferase (HAT) activity of CBP acetylates the N-terminal tails of histones such as H3 during transcriptional activation^{20–23}. One potential effect of histone H3 phosphorylation has been proposed to be the recruitment and activation of HAT^{22,23}. For example, the phosphorylation of Ser 10 in histone H3 enhances the HAT-mediated acetylation of Lys 14. Such modifications of specific residues in the N-terminal tails of histones might serve as a signal for the binding of specific co-activators to result in increased gene expression.

Because the kinase that regulates the cytokine-induced phosphorylation of histone H3 has not been identified, we addressed whether IKK- α might be involved in this process. The state of acetylation and phosphorylation of histone H3 in MEF, IKK- α ^{-/-} and IKK- β ^{-/-} cells that had been serum-starved for 16 h before treatment with TNF- α was characterized. Treatment of MEF cells with TNF- α resulted in an increased phosphorylation of Ser 10 and acetylation of Lys 14 in histone H3 (Fig. 5a). In IKK- α ^{-/-} cells there was a significantly decreased phosphorylation of Ser 10 and acetylation of Lys 14 in histone H3 both in the absence and in the presence of TNF- α in comparison with that seen in MEF cells (Fig. 5a). IKK- β ^{-/-} cells contained similar levels of histone H3 phosphorylation to that detected in MEF cells, but exhibited marked decreases in the acetylation of Lys 14 in histone H3. Treatment of each of these cells with trichostatin A, an inhibitor of histone deacetylases, resulted in similar levels of histone acetylation. These results suggested that IKK- α may play a role in phosphorylation of Ser 10, which might be important for the subsequent acetylation of Lys 14 in histone H3.

Next, kinase assays were performed *in vitro* to determine whether IKK- α could directly phosphorylate histone H3. Kinase assays performed with purified histone H3 as substrate demonstrated that wild-type IKK- α , but not the kinase-defective mutants IKK- α (SS/AA) or IKK- α (K/M), resulted in enhanced histone H3 phosphorylation (Fig. 5b). Consistent with these results was the observation from western blot analysis that the presence of IKK- α resulted in increased levels of phosphorylated Ser 10 and acetylated Lys 14 in histone H3 (Fig. 5c). ChIP assays demonstrated that IKK- α led to the increased association with the I κ B α promoter of histone H3 that was phosphorylated on Ser 10 and acetylated on Lys 14 without changing the association of total histone H3 and phosphorylated histone H1 (Fig. 5d). These results indicated that IKK- α was likely to be responsible for cytokine-induced phosphorylation and subsequent acetylation in histone H3.

Finally, we addressed the kinetics of TNF- α -mediated increases in the association of IKK- α , p65, CBP and phosphorylated and acetylated histone H3 with the I κ B α promoter (Fig. 5e). ChIP

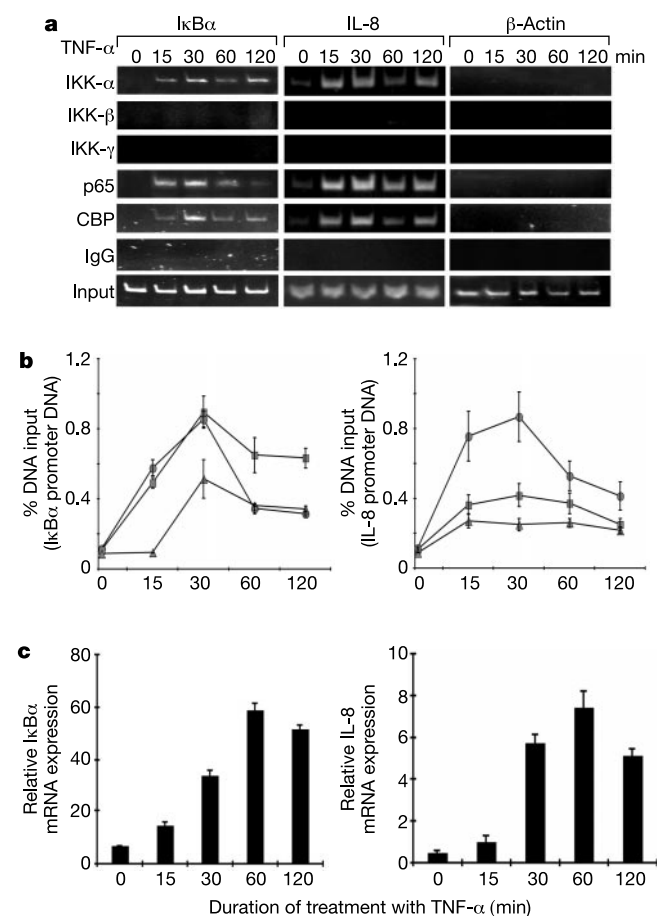


Figure 2 TNF- α treatment leads to IKK- α association with the I κ B α and IL-8 promoters. **a**, HeLa cells were treated with TNF- α , and chromatin immunoprecipitation assays were performed with the indicated antibodies. The detection of the immunoprecipitated I κ B promoter (left panel), IL-8 promoter (middle panel) or β -actin promoter (right panel) was analysed by PCR with promoter-specific primers. **b**, Quantitative real-time PCR was performed to determine the kinetics of factor binding to these promoters in comparison with input DNAs. Squares, IKK- α ; circles, p65; triangles, CBP. **c**, I κ B α and IL-8 mRNA levels were determined by quantitative real-time PCR and normalized by measuring the levels of 18S RNA.

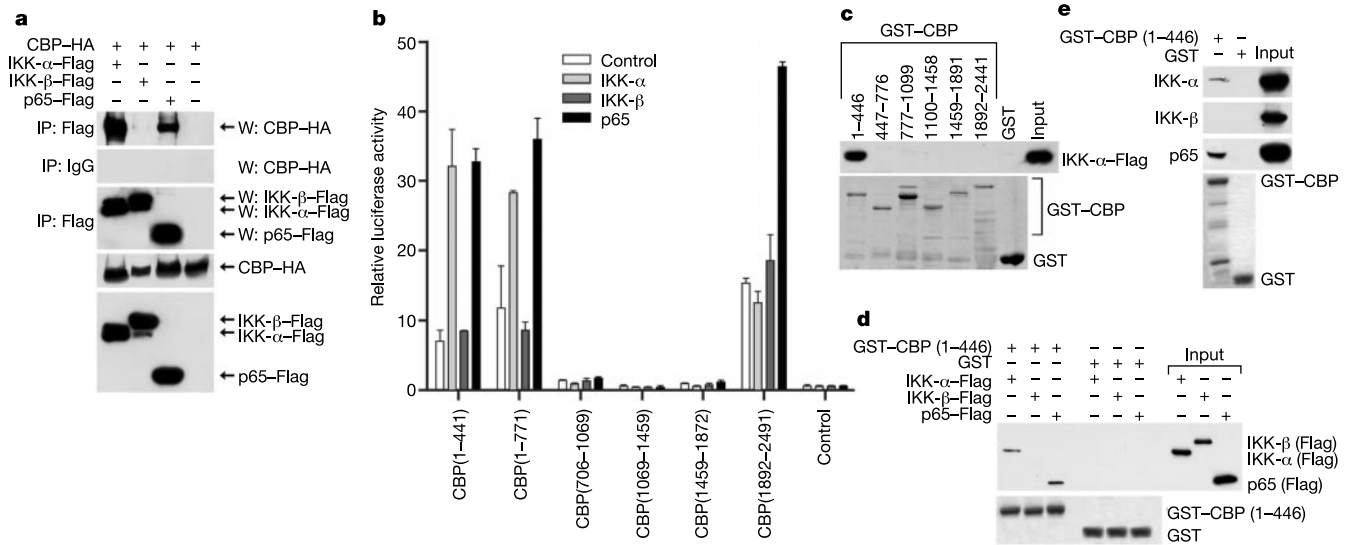


Figure 3 IKK- α interacts with the CBP transactivation domain. **a**, Flag-tagged IKK- α , IKK- β or p65 was co-transfected with haemagglutinin (HA)-tagged CBP into HEK-293 T cells. After immunoprecipitation (IP) with anti-Flag antibody, the immunoprecipitates were analysed by immunoblotting (W) with anti-HA or anti-Flag antibody as indicated. **b**, The indicated regions of CBP fused to the GAL4 DNA-binding domain were co-transfected

with a GAL4 luciferase reporter construct and IKK- α , IKK- β or p65 constructs containing the VP16 activation domain. **c-e**, GST-CBP fusion proteins were used to determine their interaction with Flag-tagged IKK- α , IKK- β or p65 in extracts prepared from transfected cells (**c** and **d**) or with endogenous proteins isolated from HeLa cells (**e**).

assays of MEF and IKK- $\alpha^{-/-}$ cells demonstrated that TNF- α treatment led to the recruitment of both p65 and CBP to the I κ B α promoter (Fig. 5e). Whereas treatment of MEF cells with TNF- α led to the enhanced promoter association of histone H3 phosphorylated on Ser 10 and acetylated on Lys 14, little or no promoter association of histone H3 modified in this manner was

found in IKK- $\alpha^{-/-}$ cells. ChIP assays of the I κ B α promoter in IKK- $\beta^{-/-}$ cells required the addition of threefold more input DNA to facilitate the detection of factor binding than in the other cell lines, owing to the decreased association of IKK- α , CBP and p65 with the I κ B α promoter (Fig. 5e). Treatment of these cells with TNF- α resulted in an increased association of histone H3

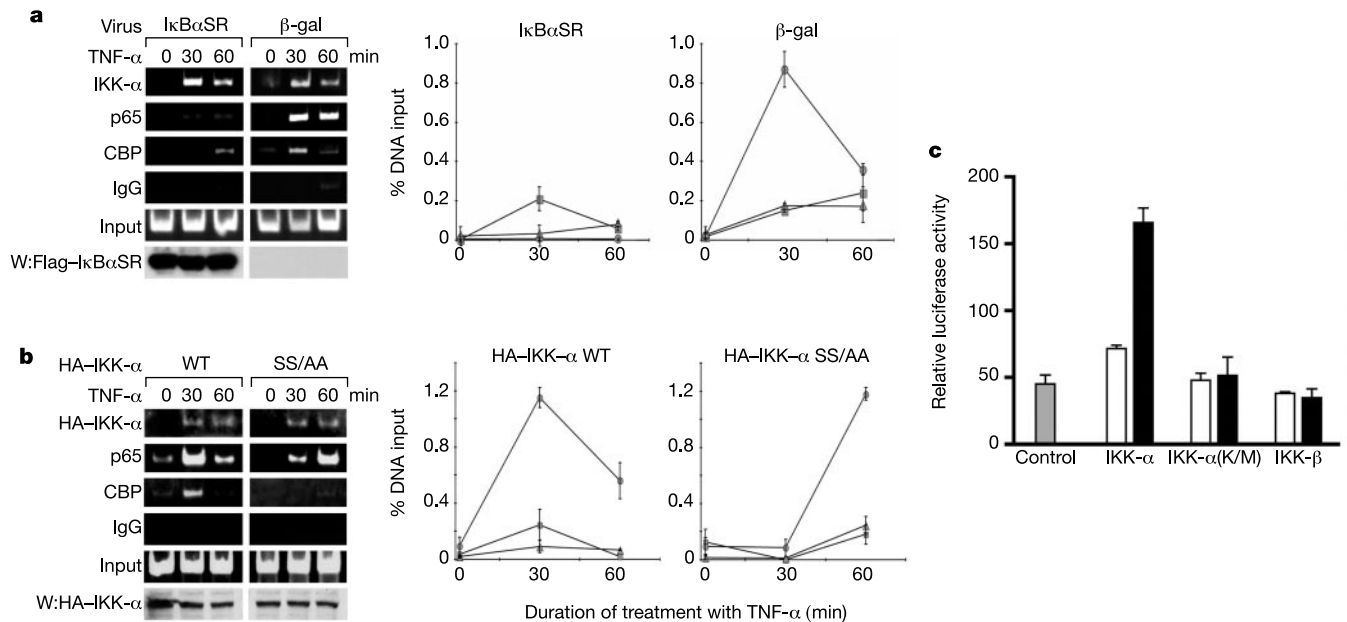


Figure 4 IKK- α promoter association is not strictly dependent on p65 binding and results in increased CBP transcriptional activity. **a**, HeLa cells infected with adenovirus encoding either the I κ B α super-repressor (I κ B α SR) or β -galactosidase (β -gal) were stimulated with TNF- α for the specified durations. The association of IKK- α , p65 and CBP with the I κ B α promoter was analysed by ChIP assays followed by PCR analysis (left panel) or quantitative real-time PCR (right panels). Squares, IKK- α ; circles, p65; triangles, CBP. **b**, After transfection of either haemagglutinin (HA)-tagged IKK- α (WT) or IKK- α (SS/AA) into HeLa

cells, ChIP assays (left panel) and quantitative real-time PCR (right panels) were performed. Squares, IKK- α ; circles, p65; triangles, CBP. Immunoblotting (W) of the epitope-tagged proteins is also indicated. **c**, An expression vector encoding a GAL4-CBP fusion protein was transfected with a GAL4 luciferase reporter construct and either 0.1 μ g (open bars) or 0.5 μ g (filled bars) of expression vectors encoding IKK- α , IKK- α (K/M) or IKK- β tagged with Myc, and CBP-dependent transcriptional activity was determined by luciferase assays.

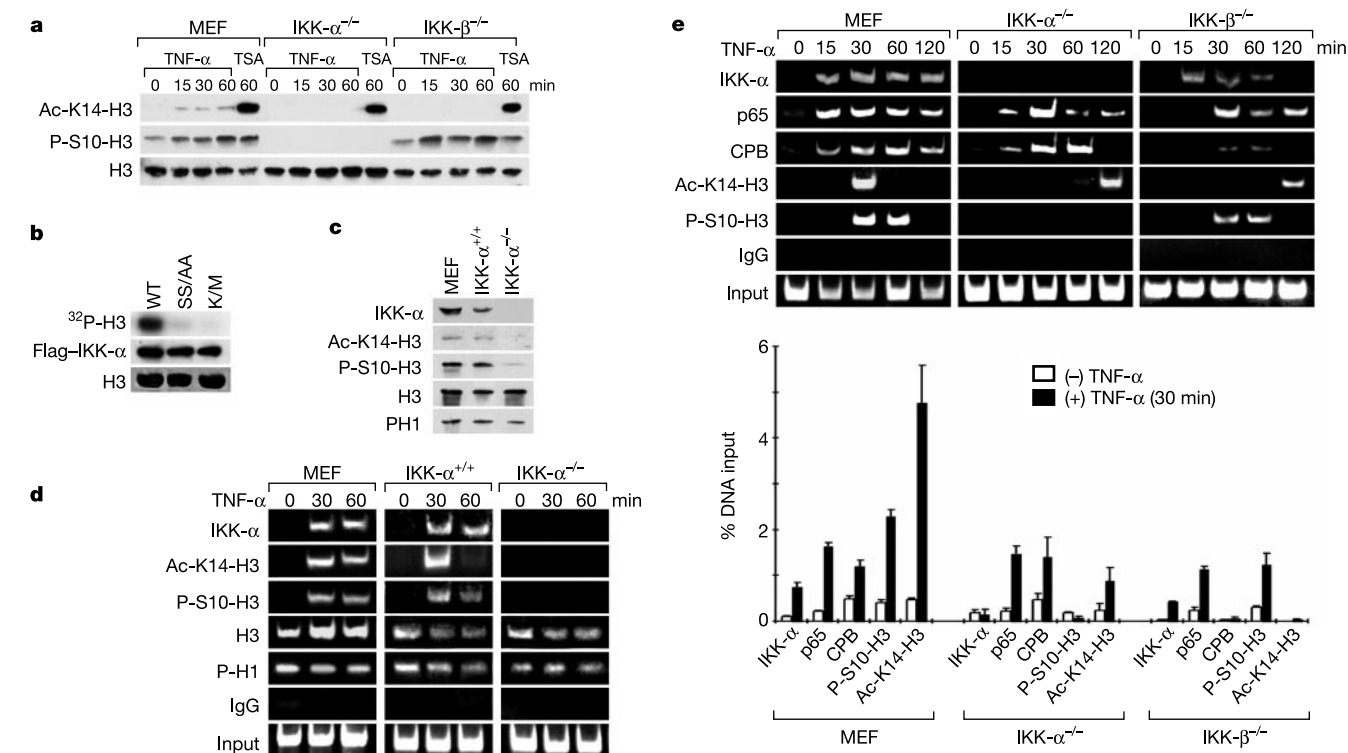


Figure 5 Requirement for IKK- α in histone H3 phosphorylation. **a**, Each of the indicated cell lines was serum-starved for 16 h, then treated with TNF- α for the specified durations or with trichostatin A (TSA) for 60 min. Acid-extracted proteins from these cells were immunoblotted with the indicated antibodies. **b**, Flag-tagged IKK- α , IKK- α (SS/AA) or IKK- α (K/M) constructs were transfected into HeLa cells and the kinase activity of these immunoprecipitated Flag-tagged proteins was determined with purified histone H3 as substrate. **c**, Extracts prepared from IKK- $\alpha^{-/-}$, IKK- $\alpha^{+/+}$ or MEF cells were analysed for

IKK- α expression and the acid-extracted proteins from these cells were immunoblotted with the indicated antibodies. **d**, The association of these factors with the I κ B α promoter was also determined by ChIP assays (lower panel). **e**, MEF, IKK- $\alpha^{-/-}$ and IKK- $\beta^{-/-}$ were treated with TNF- α for the specified durations and ChIP assays were performed and analysed by PCR (top panel) or quantitative real-time PCR at 0 and 30 min after treatment with TNF- α (lower panel).

phosphorylated on Ser 10, but a markedly reduced association of both CBP and histone H3 acetylated on Lys 14 compared with that seen in MEF cells (Fig. 5e). These studies suggested that the binding of both IKK- α and CBP to the I κ B α promoter was important for the association of histone H3 phosphorylated on Ser 10 and acetylated on Lys 14.

Phosphorylation of histone H3 is associated with several distinct biological processes including the mitogen-induced and cytokine-induced activation of specific genes and the compaction of chromosomes at the onset of mitosis^{20,21}. For example, the Aurora kinases have been shown to be important in the phosphorylation of Ser 10 in histone H3 during mitosis²¹. Stimulation of both the extracellular signal-regulated protein kinase (ERK) pathway and the stress-activated p38 pathway also result in an increase in phosphorylation of histone H3 (Ser 10) during the activation of immediate-early gene expression^{20,21}. The Rsk-2 (ref. 24) and Msk-1 (ref. 25) kinases have been shown to be important in this process. Previous studies of the I κ B α promoter indicated that an uncharacterized kinase distinct from Rsk-2 and Msk-1 could phosphorylate histone H3 during cytokine-mediated activation of this promoter²⁶. Our analysis and that of a companion paper²⁷ indicate that IKK- α , but not IKK- β , is likely to be the kinase that is critical in facilitating the rapid cytokine-induced expression of NF- κ B-regulated genes.

IKK- α recruitment to NF- κ B regulated promoters is probably mediated by multiple promoter-bound factors including components of the basal transcription complex, co-activators such as CBP, and perhaps chromatin itself. The cytokine-induced phosphorylation by IKK- α of Ser 10 in histone H3 seems to be especially important for the subsequent acetylation of Lys 14 by CBP. This

result is consistent with previous studies indicating that phosphorylation of Ser 10 precedes the subsequent acetylation of Lys 14 in histone H3 and that the HAT activity for H3 phosphorylated on Ser 10 is enhanced in comparison with that for unmodified H3 (refs 22, 23). In addition to its previously described role on the processing of p100 to p52 (ref. 28), our analysis shows that IKK- α is important in regulating a critical step required for NF- κ B-dependent gene expression. The studies presented here extend our knowledge of the multiple functions of the I κ B kinases in activating the NF- κ B pathway. □

Methods

Cells and reagents

Mouse embryo fibroblasts (MEFs) were a gift from Xiaodong Wong. IKK- $\alpha^{-/-}$ and IKK- $\beta^{-/-}$ cells were provided by Inder M. Verma. A Moloney-based retrovirus expressing (MEQKLISEEDLN) (Myc)-tagged IKK- α was used with selection by hygromycin to stably express IKK- α in IKK- $\alpha^{-/-}$ cells to generate IKK- $\alpha^{+/+}$ cells. Antibodies directed against IKK- α / β , I κ B α , p65 and CBP were obtained from Santa Cruz Biotechnology. IKK- α -specific antibody was from BD/Pharmingen and Oncogene and IKK- β -specific antibody was obtained from Cell Signaling Technology. Antibodies directed against acetyl-histone H3 (Lys 14), phospho-histone H3 (Ser 10) and phospho-histone H1 were obtained from Upstate Biotechnology. Antibody against histone H3 was obtained from Cell Signaling Technology. TNF- α (Roche Molecular Biochemicals) was used at a final concentration of 10 ng ml⁻¹, and trichostatin A (Sigma) was used at 500 ng ml⁻¹.

Quantitative real-time PCR

Total RNA was prepared from HeLa, MEF, IKK- $\alpha^{-/-}$, IKK- $\alpha^{+/+}$ or IKK- $\beta^{-/-}$ cells with the RNeasy kit (Qiagen) and treated with the DNA-free kit (Ambion) to remove residual genomic DNA. For real-time PCR, the cDNA was prepared with oligo(dT) and random primers (Invitrogen) and analysed in triplicate with the SYBR GreenMaster Mix (Applied Biosystems) for 15 min at 95 °C for initial denaturing, followed by 40 cycles of 95 °C for 30 s and 60 °C for 30 s in the ABI Sequence Detection System. The oligonucleotide

primers used to analyse I κ B α transcripts (+2180 to +2378) contained the sequences 5'-GATCCGCCAGGTGAAGGG-3' and 5'-GCAATTCTGGCTGGTTGG-3', the primers used to analyse IL-8 transcripts (+1 to +239) contained the sequences 5'-ATGACTTCCAGCTGGCCGT-3' and 5'-TTACATAATTCTGTGTGGC-3', and the primers used to analyse the 18S ribosomal RNA contained the sequences 5'-AGGAATTGACGGAAGGGCAC-3' and 5'-GGACATCTAAGGGCATCACA-3'.

ChIP assays

ChIP assays were performed with a previously described protocol (Upstate Biotechnology). In brief, chromatin from crosslinked cells was sheared by sonication (three times, 15 s each; one-third power) and incubated overnight with specific antibody followed by incubation with protein G-Sepharose saturated with salmon sperm DNA. Precipitated DNAs were analysed by quantitative PCR (34 cycles) with a Taq PCR Master mix kit (Qiagen) and primers for either the human 5'-GACGACCCCAATTCAAATCG-3' and 5'-TCAGGCTCGGGAATTTC-3' or murine 5'-GGACCCCAACCAAAATCG-3' and 5'-TCAGGCGGGGGAATTTC-3' I κ B α promoters (-316 to -15), together with the human IL-8 promoter (-121 to +61) 5'-GGGCCATCAGTTGCAATC-3' and 5'-TTCCTTCGGTGGTTTCTTC-3' and the human β -actin promoter (-980 to -915) 5'-TGCACTGTGCGGCGAAGC-3' and 5'-TCGAGCCATAAAAGGCAA-3'. Quantitative real-time PCR was performed in triplicate to determine the association of IKK- α , p65 and CBP with the I κ B α and IL-8 promoters by using 500 nM of the above oligonucleotide primers and input DNA standards diluted in threefold increments from 10% to 0.01% with SYBR Green Master Mix and the ABI Prism 7700 Sequence Detection System.

In vitro interaction assay

Fragments of the CBP coding sequence were cloned into pGEX vector (Pharmacia). Purified GST-CBP fusion proteins were immobilized to glutathione-agarose and incubated overnight with cell lysates (100 μ g protein). After extensive washing with cold PBS, the protein complexes were analysed by immunoblotting.

IKK- α kinase assay

Total cell lysates (100 μ g protein) prepared from cells transfected with expression vectors encoding Flag-tagged IKK- α , IKK- α (K/M) or IKK- α (S/A) were incubated for 1 h with anti-Flag antibody (Sigma) and with protein A-agarose for a further 1 h. After extensive washing of the immunoprecipitates, kinase assays were performed as described²⁹ with 5 μ g of histone H3 (Sigma) as a substrate.

Received 4 February; accepted 21 March 2003; doi:10.1038/nature01576.

- Baldwin, A. S. The NF- κ B and I κ B proteins: new discoveries and insights. *Annu. Rev. Immunol.* **14**, 649–681 (1996).
- Ghosh, S., May, M. J. & Kopp, E. B. NF- κ B and Rel proteins: evolutionarily conserved mediators of immune responses. *Annu. Rev. Immunol.* **16**, 225–260 (1998).
- Silverman, N. & Maniatis, T. NF- κ B signaling pathways in mammalian and insect innate immunity. *Genes Dev.* **15**, 2321–2342 (2001).
- Zandi, E. & Karin, M. Bridging the gap: composition, regulation, and physiological function of the I κ B kinase complex. *Mol. Cell. Biol.* **19**, 4547–4551 (1999).
- Takeda, K. *et al.* Limb and skin abnormalities in mice lacking IKK α . *Science* **284**, 313–316 (1999).
- Hu, Y. *et al.* Abnormal morphogenesis but intact IKK activation in mice lacking the IKK α subunit of I κ B kinase. *Science* **284**, 316–320 (1999).
- Li, Q. *et al.* IKK1-deficient mice exhibit abnormal development of skin and skeleton. *Genes Dev.* **13**, 1322–1328 (1999).
- Sizemore, N., Lerner, N., Dombrowski, N., Sakurai, H. & Stark, G. R. Distinct roles of the I κ B kinase α and β subunits in liberating nuclear factor κ B (NF- κ B) from I κ B and in phosphorylating the p65 subunit of NF- κ B. *J. Biol. Chem.* **277**, 3863–3869 (2002).
- Li, X. *et al.* IKK α , IKK β and NEMO/IKK γ are each required for the NF- κ B mediated inflammatory response program. *J. Biol. Chem.* **277**, 45129–45140 (2002).
- Birbach, A. *et al.* Signaling molecules of the NF- κ B pathway shuttle constitutively between cytoplasm and nucleus. *J. Biol. Chem.* **277**, 10842–10851 (2002).
- Jankecht, R. & Hunter, T. Transcription. A growing coactivator network. *Nature* **383**, 22–23 (1996).
- Goodman, R. H. & Smolik, S. CBP/p300 in cell growth, transformation, and development. *Genes Dev.* **14**, 1553–1577 (2000).
- Zhong, H., Voll, R. E. & Ghosh, S. Phosphorylation of NF- κ B p65 by PKA stimulates transcriptional activity by promoting a novel bivalent interaction with the coactivator CBP/p300. *Mol. Cell* **1**, 661–671 (1998).
- Zhong, H., May, M. J., Jimi, E. & Ghosh, S. The phosphorylation status of nuclear NF- κ B determines its association with CBP/p300 or HDAC-1. *Mol. Cell* **9**, 625–636 (2002).
- Sakurai, H., Chiba, H., Miyoshi, H., Sugita, T. & Toriumi, W. I κ B kinases phosphorylate NF- κ B p65 subunit on serine 536 in the transactivation domain. *J. Biol. Chem.* **274**, 30353–30356 (1999).
- Madrid, L. V., Mayo, M. W., Reuther, J. Y. & Baldwin, A. S. Jr Akt stimulates the transactivation potential of the RelA/p65 subunit of NF- κ B through utilization of the I κ B kinase and activation of the mitogen-activated protein kinase p38. *J. Biol. Chem.* **276**, 18934–18940 (2001).
- Gerritsen, M. E. *et al.* CREB-binding protein/p300 are transcriptional coactivators of p65. *Proc. Natl Acad. Sci. USA* **94**, 2927–2932 (1997).
- Wang, C. Y., Mayo, M. W. & Baldwin, A. S. J. TNF α - and cancer therapy-induced apoptosis: potentiation by inhibition of NF- κ B. *Science* **274**, 784–787 (1996).
- Van Antwerp, D. J., Martin, S. J., Kafri, T., Green, D. R. & Verma, I. M. Suppression of TNF α -induced apoptosis by NF- κ B. *Science* **274**, 787–789 (1996).
- Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
- Cheung, P., Allis, C. D. & Sassone-Corsi, P. Signaling to chromatin through histone modifications. *Cell* **103**, 263–271 (2000).
- Cheung, P. *et al.* Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. *Mol. Cell* **5**, 905–915 (2000).
- Lo, W.-S. *et al.* Phosphorylation of serine 10 in histone H3 is functionally linked *in vitro* and *in vivo* to Gcn5-mediated acetylation at lysine 14. *Mol. Cell* **5**, 917–926 (2000).

- Sassone-Corsi, P. *et al.* Requirement of Rsk-2 for epidermal growth factor-activated phosphorylation of histone H3. *Science* **285**, 886–891 (1999).
- Thomson, S. *et al.* The nucleosomal response associated with immediate-early gene induction is mediated via alternative MAP kinase cascades: MSK1 as a potential histone H3/HMG-14 kinase. *EMBO J.* **18**, 4779–4793 (1999).
- Saccani, S., Pantano, S. & Natoli, G. p38-dependent marking of inflammatory genes for increased NF- κ B recruitment. *Nature Immunol.* **3**, 69–75 (2002).
- Anest, V. *et al.* A nucleosomal function for I κ B kinase- α in NF- κ B-dependent gene expression. *Nature* **423**, 659–663 (2003).
- Senfleben, U. *et al.* Activation by IKK α of a second, evolutionary conserved, NF- κ B signaling pathway. *Science* **293**, 1495–1499 (2001).
- Yamamoto, Y. *et al.* IKK γ /NEMO facilitates the recruitment of the I κ B proteins into the I κ B kinase complex. *J. Biol. Chem.* **276**, 36327–36336 (2001).

Acknowledgements We thank J. Guo for assistance, A. Herrera and M. Singh for preparation of the figures, J. Darrah for preparing the manuscript, and T. Collins for GAL4-CBP constructs.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.B.G. (gaynor_richard@lilly.com).

A nucleosomal function for I κ B kinase- α in NF- κ B-dependent gene expression

Vasiliki Anest^{*†}, Julie L. Hanson^{*†}, Patricia C. Cogswell^{*}, Kris A. Steinbrecher^{*}, Brian D. Strahl^{*‡§} & Albert S. Baldwin^{*†§}

^{*} Lineberger Comprehensive Cancer Center,
[†] Curriculum in Genetics and Molecular Biology,
[‡] Department of Biochemistry and Biophysics, and
[§] Department of Biology, University of North Carolina, Chapel Hill, North Carolina 27599, USA

NF- κ B is a principal transcriptional regulator of diverse cytokine-mediated processes and is tightly controlled by the I κ B kinase complex (IKK- α / β / γ). IKK- β and IKK- γ are critical for cytokine-induced NF- κ B function, whereas IKK- α is thought to be involved in other regulatory pathways^{1–4}. However, recent data suggest a role for IKK- α in NF- κ B-dependent gene expression in response to cytokine treatment^{1,5–7}. Here we demonstrate nuclear accumulation of IKK- α after cytokine exposure, suggesting a nuclear function for this protein. Consistent with this, chromatin immunoprecipitation (ChIP) assays reveal that IKK- α was recruited to the promoter regions of NF- κ B-regulated genes on stimulation with tumour-necrosis factor- α . Notably, NF- κ B-regulated gene expression is suppressed by the loss of IKK- α and this correlates with a complete loss of gene-specific phosphorylation of histone H3 on serine 10, a modification previously associated with positive gene expression. Furthermore, we show that IKK- α can directly phosphorylate histone H3 *in vitro*, suggesting a new substrate for this kinase. We propose that IKK- α is an essential regulator of NF- κ B-dependent gene expression through control of promoter-associated histone phosphorylation after cytokine exposure. These findings provide additional insight into the role of the IKK complex in NF- κ B-regulated gene expression.

Characteristic cytokine-mediated activation of the NF- κ B pathway involves IKK- β -directed phosphorylation of and subsequent degradation of inhibitors of NF- κ B (I κ Bs), resulting in rapid nuclear accumulation of NF- κ B subunits^{1–3}. It has been found that IKK- α , but not IKK- β , constitutively shuttles between the cytoplasm and the nucleus, suggesting a nuclear function for this IKK subunit⁸. To address whether the subcellular localization of IKK- α changes on

cytokine exposure, nuclear and cytoplasmic extracts from mouse embryonic fibroblasts (MEFs) stimulated with tumour-necrosis factor- α (TNF- α) were immunoblotted with antibodies against p65 and the different IKK subunits. Notably, stimulation of MEFs with TNF- α results in marked nuclear accumulation of IKK- α (Fig. 1a). In contrast to IKK- α , both IKK- β and IKK- γ are present in the nucleus and the cytoplasm, and exhibit no appreciable changes on TNF- α induction (Fig. 1a). We demonstrated the purity of nuclear and cytoplasmic fractions with antibodies specific for p105, an NF- κ B precursor known to be cytoplasmic, and for TFIIB, a known nuclear protein (Fig. 1a). Immunofluorescence staining with an anti-IKK- α antibody confirms the western analysis. IKK- α is predominantly cytoplasmic with low but detectable nuclear levels in unstimulated wild-type MEFs, whereas TNF- α -induced cells exhibit higher levels of nuclear IKK- α (Fig. 1b, left panel). Loss of immunoreactivity in IKK- $\alpha^{-/-}$ cells confirms antibody specificity (Supplementary Information). Consistent with previous reports, TNF- α treatment induces nuclear translocation of p65 (Fig. 1b, right panel). Although the kinetics of induced IKK- α nuclear accumulation are similar to those of p65, TNF- α induction leads to nuclear accumulation of IKK- α in p65 $^{-/-}$ cells (data not shown), indicating that IKK- α does not require p65 for nuclear entry. These data suggest that IKK- α may have a nuclear function distinct from the canonical NF- κ B activation pathway.

Consistent with previous reports¹, we observed normal TNF- α -induced I κ B α degradation and induced p65 nuclear translocation in MEFs lacking IKK- α (data not shown). However, I κ B α resynthesis is delayed in IKK- $\alpha^{-/-}$ cells compared with wild-type MEFs (Fig.

1c). The defect in resynthesis is consistent with a previous report implicating a role for IKK- α in TNF- α -induced I κ B α gene expression⁵. Accordingly, we performed real-time polymerase chain reaction (PCR) analysis to examine the expression profile of the NF- κ B-regulated I κ B α gene (*Nfkb1a*) in IKK- $\alpha^{-/-}$, IKK- $\beta^{-/-}$ and wild-type MEFs. TNF- α rapidly induces I κ B α gene expression in wild-type MEFs (Fig. 2a), whereas TNF- α -induced I κ B α gene expression is significantly decreased in both IKK- $\alpha^{-/-}$ and IKK- $\beta^{-/-}$ cells. We next examined the kinetics of TNF- α -induced interleukin-6 (IL-6) gene expression by northern analysis. Both IKK- $\alpha^{-/-}$ and IKK- $\beta^{-/-}$ cells are deficient in IL-6 gene (*Il6*) activation at all time points after TNF- α -induction (Fig. 2c). Previous reports have shown that loss of IKK- α and IKK- β results in altered transcriptional activation of I κ B α , IL-6 and other NF- κ B-dependent genes⁵⁻⁷. Thus, our results support a requirement for IKK- α and IKK- β in optimal I κ B α and IL-6 gene expression.

The molecular mechanism(s) through which IKK- α regulates the expression of specific cytokine-induced NF- κ B-dependent genes has not been determined previously. To investigate a nuclear role for IKK- α , we performed ChIP assays with antibodies directed against p65, IKK- α and IKK- β in TNF- α -stimulated wild-type MEFs. Furthermore, we used phosphorylated and acetylated histone H3 antibodies to examine histone H3 modifications that are known to correlate with active gene expression at the I κ B α promoter⁹. The results reveal a rapid recruitment of p65 to the I κ B α promoter in response to TNF- α (Fig. 2b, left panel). Of note, both IKK- α and IKK- β recruitment are observed after TNF- α -induction, whereas we detect a low level of IKK- α recruitment in unstimulated cells.

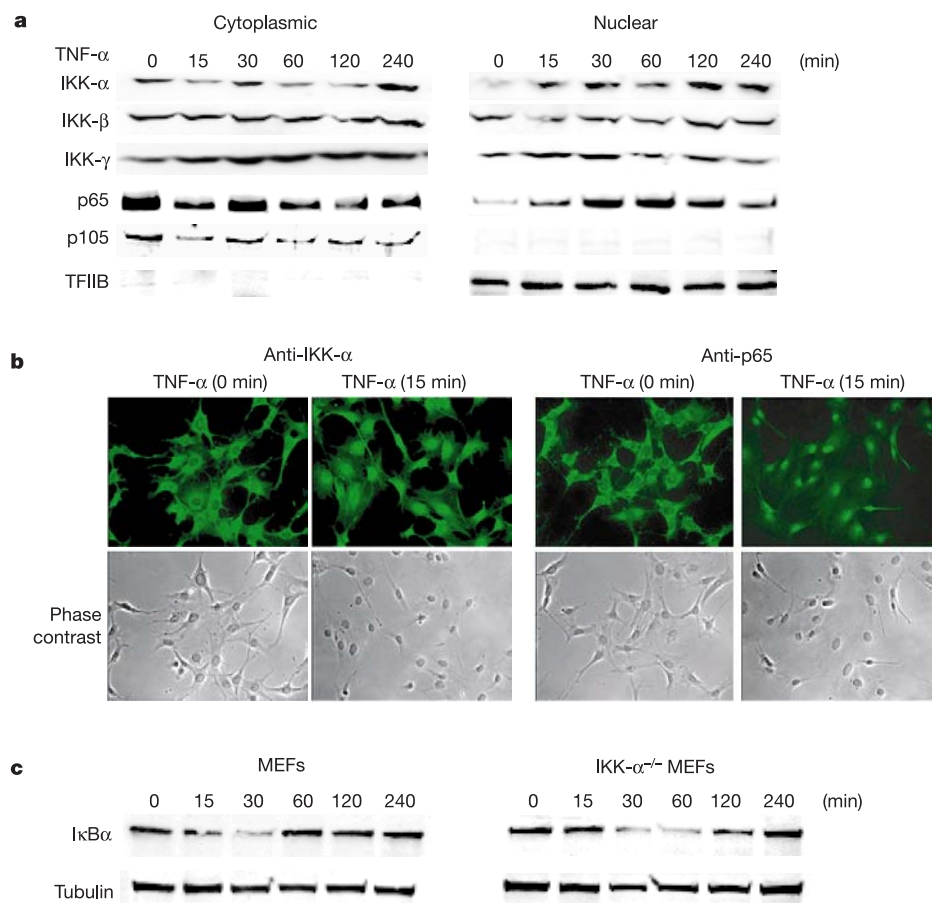


Figure 1 TNF- α -induced nuclear accumulation of endogenous IKK- α . **a**, MEFs were stimulated with TNF- α (10 ng ml $^{-1}$) and fractionated into nuclear and cytoplasmic fractions, and then analysed directly by western blotting with IKK- α , IKK- β , IKK- γ and p65 antibodies. Antibodies specific for cytoplasmic p105/p50 and nuclear TFIIB demonstrate the purity of these fractions. **b**, MEFs were treated for 15 min with TNF- α as indicated,

and were processed for indirect immunofluorescence and stained with IKK- α or p65 antibodies. **c**, Kinetics of TNF- α -induced I κ B α degradation in IKK wild-type and IKK- $\alpha^{-/-}$ MEFs. Cells were treated with TNF- α for the indicated times and western analysis was performed with anti-I κ B α antibody.

The kinetics of IKK- α recruitment and histone H3 Ser 10 phosphorylation correlate in uninduced and TNF- α -induced cells. Consistent with previous reports and with evidence of basal I κ B α gene expression⁹⁻¹¹, we observed constitutive levels of acetylated histone H3 at this promoter (Fig. 2b, left panel). Enrichment of promoter sequences is not detected at the β -actin promoter or when

the antibody is omitted from the immunoprecipitation reaction, confirming the specificity of the ChIP assay (data not shown). These results demonstrate a TNF- α -inducible association of IKK- α and IKK- β with the I κ B α promoter, and suggest a new function for these kinases in NF- κ B-dependent gene regulation.

To address whether IKK- α has a unique nuclear function in modulating gene expression, we examined the recruitment profile of p65 and the status of histone modifications at the I κ B α promoter in IKK- α ^{-/-} cells following TNF- α treatment. The kinetics of p65 recruitment in these cells is similar to those of wild-type MEFs, whereas IKK- β recruitment is delayed (Fig. 2b, right panel), indicating that p65 and IKK- β association with the I κ B α promoter does not require IKK- α . The levels of TNF- α -induced p65 recruitment are reduced in IKK- α ^{-/-} cells. Notably TNF- α -induced H3 Ser 10 phosphorylation levels are completely abolished in the IKK- α ^{-/-} cells. Basal levels of H3 acetylation are decreased in the IKK- α ^{-/-} cells but normal levels are restored after TNF- α -induction. These data indicate that IKK- α is required for Ser 10 phosphorylation of histone H3 at the I κ B α promoter. The requirement for IKK- α in controlling H3 Ser 10 phosphorylation is also observed on another immediately accessible NF- κ B-regulated gene, *Mip2* (also known as *Cxcl2*; data not shown).

Next, we investigated recruitment of p65 and IKK- α , and the status of H3 Ser 10 phosphorylation at the I κ B α promoter, using IKK- β ^{-/-} MEFs. We detected reduced and transient levels of promoter-associated p65 and IKK- α after stimulation with TNF- α . The onset of Ser 10 phosphorylation matches the pattern of IKK- α recruitment in these cells after stimulation with TNF- α (Fig. 2b, lower panel). These data show that the loss of IKK- β does not affect Ser 10 phosphorylation at the I κ B α promoter.

To examine further a requirement for IKK- α in regulating Ser 10 phosphorylation of histone H3, we evaluated the IL-6 gene, an NF- κ B-regulated gene that requires prior chromatin modifications to ensure accessibility to NF- κ B¹¹. In wild-type MEFs, p65, IKK- α and IKK- β all show a similar profile of promoter recruitment occurring 2 h after stimulation with TNF- α (Fig. 2d, left panel). Similar to the results obtained with the I κ B α promoter, the presence of IKK- α directly correlates with the onset of H3 Ser 10 phosphorylation. Promoter-associated IKK- α levels decrease 4 h after treatment whereas Ser 10 levels remain detectable, suggesting that either additional H3 kinases may be involved in the maintenance of Ser 10 levels at the IL-6 promoter or that Ser 10 phosphorylation levels remain stable after transient recruitment of IKK- α . Notably, we did not detect TNF- α -induced Ser 10 phosphorylation in IKK- α ^{-/-} cells at the IL-6 promoter (Fig. 2d, right panel). p65 recruitment in IKK- α ^{-/-} cells exhibits similar kinetics to wild-type cells, although overall levels appear to be lower. We did not detect IKK- β at this promoter in IKK- α ^{-/-} cells, suggesting a role for IKK- α in recruiting IKK- β to the IL-6 promoter. Collectively, these results indicate a requirement for IKK- α in mediating H3 Ser 10 phosphorylation at different NF- κ B-dependent promoters.

The timing of TNF- α -induced expression of IL-6 messenger RNA does not coincide with recruitment of p65 to the IL-6 promoter (Fig. 2c, d). We examined IL-6 mRNA levels in wild-type and p65^{-/-} MEFs to determine whether expression of IL-6 is strictly p65-dependent. Northern blotting experiments revealed decreased IL-6 expression in p65^{-/-} cells, although low-level induction of IL-6 by TNF- α is observed (data not shown). This is consistent with p65 being necessary for efficient IL-6 transcription but suggests that other transcription factors, probably additional NF- κ B subunits (J.L.H., unpublished observations), can initiate IL-6 transcription at a minimal level after TNF- α stimulation.

The co-occupancy of p65 and IKK- α at both I κ B α and IL-6 promoters in wild-type MEFs stimulated with TNF- α led us to examine whether p65 is required for the recruitment of IKK- α at these promoters. In p65^{-/-} MEFs, neither IKK- α recruitment nor Ser 10 phosphorylation is detected, indicating that p65 activation

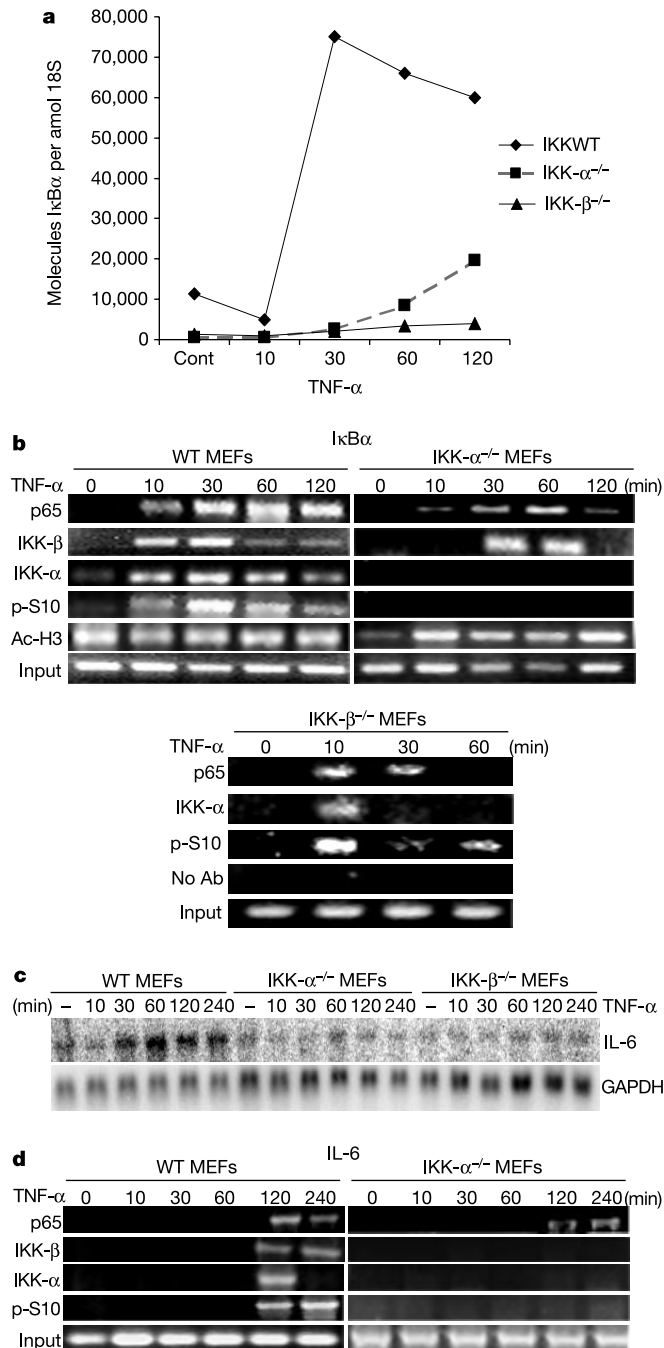


Figure 2 Promoter-associated IKK- α is essential for TNF- α -induced NF- κ B-dependent gene expression and histone H3 phosphorylation. **a**, **c**, IKK- α ^{-/-} and IKK- β ^{-/-} cells are defective in TNF- α -induced I κ B α and IL-6 gene expression. IKK wild-type (WT), IKK- α ^{-/-} and IKK- β ^{-/-} MEFs were stimulated with TNF- α and RNA levels were measured by real-time PCR for I κ B α (**a**) and northern blot analysis for IL-6 gene expression (**c**). **b**, **d**, Chromatin immunoprecipitation assays were performed on TNF- α -induced IKK WT (left panel), IKK- α ^{-/-} (right panel), or IKK- β ^{-/-} MEFs (bottom panel) with p65, IKK- α , IKK- β , H3 Ser 10 phospho-specific or acetyl-H3 antibodies (AS). Associated DNA was analysed by PCR using I κ B α (**b**) or IL-6 promoter-specific primers (**d**). These results are representative of three independent experiments.

complexes are required for IKK- α association with NF- κ B-regulated promoters (data not shown). Further investigation is needed to clarify whether p65 binding may function to alter the chromatin environment for IKK- α recruitment and subsequent Ser 10 phosphorylation of histone H3, or whether p65 directly recruits IKK- α in this response.

The role of IKK- α in controlling promoter-specific Ser 10 phosphorylation may be indirect through the regulation of other H3 kinases, or alternatively, IKK- α may directly phosphorylate histone H3. To address this latter point, we tested free core histones as substrates for recombinant IKK- α by *in vitro* kinase assays. IKK- α efficiently phosphorylates histone H3 *in vitro* in a dose-dependent fashion (Fig. 3a). Furthermore, H3 phosphorylation catalysed by IKK- α is comparable to levels obtained with known H3 kinases, MSK1 and RSK2 (refs 12, 13; Fig. 3a). Site-specificity for histone H3 phosphorylation by IKK- α was confirmed by protein immunoblotting with a well-characterized anti-H3 (Ser 10) phospho-specific antibody (Fig. 3b). IKK- β fails to phosphorylate H3 on free core histones after incubation with [γ - 32 P]ATP *in vitro* or using the Ser 10 phospho-specific antibody (Fig. 3a, b). As a positive control for IKK activity, phosphorylation of glutathione S-transferase (GST)-IkB α fusion by IKK- α and IKK- β is shown (Fig. 3b, lower panel). The analysis of the histone H3 primary sequence does not reveal the IKK phosphorylation consensus sequence. These results demonstrate that IKK- α directly phosphorylates histone H3 protein on Ser 10 *in vitro*.

Our data demonstrate TNF- α -inducible IKK- α recruitment and subsequent Ser 10 phosphorylation on specific NF- κ B-regulated promoters. To address the potential involvement of IKK- α in mechanisms associated with global levels of histone H3 phosphorylation, we took two approaches. First, we acid-extracted core histones from wild-type, IKK- α ^{-/-} and IKK- β ^{-/-} MEFs and

examined H3 Ser 10 phosphorylation by western blotting. In wild-type and IKK- β ^{-/-} MEFs, TNF- α treatment leads to increased Ser 10 phosphorylation (Fig. 4). However, both basal and TNF- α -induced levels of Ser 10 phosphorylation are significantly reduced in IKK- α ^{-/-} cells (Fig. 4) but are restored by the introduction of stably expressed IKK- α (data not shown). Previous reports suggest that impaired recognition of phosphorylated Ser 10 by phospho-H3 antibodies may occur when adjacent lysine groups are acetylated^{14,15}. Therefore, we tested an antibody against dimodified H3 to examine H3 phosphorylated on Ser 10 and acetylated on Lys 14. Levels of phosphorylated and acetylated H3 increase in response to TNF- α stimulation, with kinetics similar to that of Ser 10 phosphorylation in wild-type and IKK- β ^{-/-} MEFs (Fig. 4). However, levels of phosphorylated and acetylated H3 remain reduced in the IKK- α ^{-/-} MEFs. This suggests that the reduced affinity for Ser 10 phosphorylation in IKK- α ^{-/-} MEFs is not due to antibody occlusion by adjacent acetylation on Lys 14. In addition, global levels of acetylated H3 are largely unaffected in IKK- α ^{-/-} MEFs (data not shown). Second, immunofluorescence staining of asynchronous wild-type and IKK- α ^{-/-} cells with the same phospho-specific antibody demonstrates similar H3 Ser 10 phosphorylation levels in both cell types (data not shown). However, this probably represents mitotic-associated Ser 10 phosphorylation¹⁶, not transcriptionally associated histone phosphorylation. Thus, although it is clear that IKK- α controls promoter-associated Ser 10 phosphorylation on a set of NF- κ B-regulated genes, current investigation is aimed towards examining the role of IKK- α in modulating global levels of histone H3 phosphorylation.

Although a role for IKK- α in controlling NF- κ B-dependent gene expression has been suggested previously, a mechanism whereby IKK- α might facilitate such a response has remained unclear. Our results suggest that one role for IKK- α in controlling gene expression is through an unexpected nuclear mechanism involving histone H3 phosphorylation. Furthermore, the data indicate that IKK- α directly phosphorylates histone H3 under TNF- α stimulation. However, we cannot rule out the possibility that IKK- α is involved in the regulation of other known H3 kinases. In this regard, we observe no TNF- α -induced changes in MSK1 or RSK2 protein levels in wild-type or IKK- α ^{-/-} MEFs (data not shown). Also, MSK1 activity is minimally induced by TNF- α stimulation in wild-type or IKK- α ^{-/-} MEFs using core histones as a substrate (data not shown). Therefore, our results suggest that IKK- α -mediated histone phosphorylation may provide one nucleosomal component in the overall mechanism required for optimal gene expression. Future experiments will be necessary to determine whether IKK- α -mediated Ser 10 phosphorylation regulates other histone modifications associated with positive gene expression, such as acetylation of Lys 14 on histone H3 as predicted by the histone code hypothesis^{15,17–20}.

Our observation that IKK- β is recruited to NF- κ B-dependent promoters and is not required for histone phosphorylation (Fig. 2b, lower panel) indicates that IKK- α does not require IKK- β to control histone phosphorylation, and suggests a distinct chromatin-associated role for IKK- β in controlling NF- κ B-dependent gene

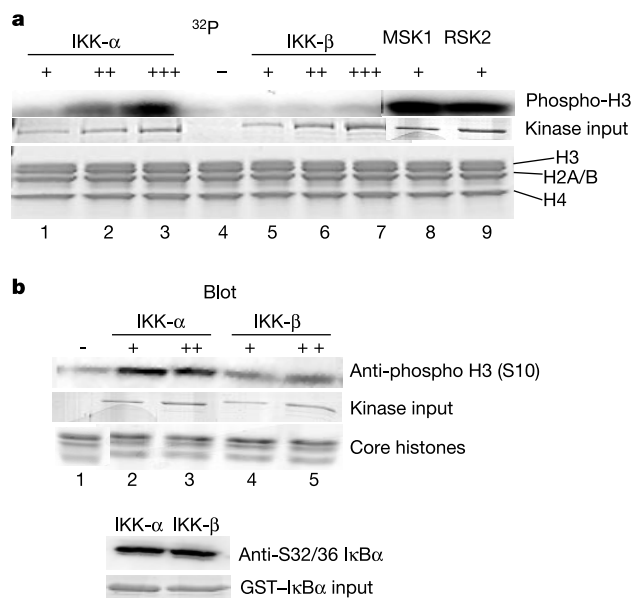


Figure 3 IKK- α directly phosphorylates histone H3 *in vitro*. **a**, Free core histones were incubated with increasing amounts of IKK- α or IKK- β (0.5, 1 and 2 μ g, respectively) and [γ - 32 P]ATP *in vitro* to analyse the dose response for histone H3 phosphorylation. For comparison, the activities of the H3 kinases (MSK1 (20 U) and RSK2 (70 U)) towards the histones were included. Reactions were resolved by SDS-PAGE (15%) and analysed by autoradiography. Coomassie staining shows histones that were incubated with either IKK- α (lanes 1–3), no kinase (lane 4), IKK- β (lanes 5–7), MSK1 (lane 8) and RSK2 (lane 9). **b**, *In vitro* kinase assays were performed essentially as described in **a**, except that they were carried out in the absence (lane 1) or presence of IKK- α (lane 2, 3) and IKK- β (lanes 3, 4), and without radioactive ATP. Western blotting with anti-phospho-specific Ser 10 antibody indicates that IKK- α directly phosphorylates H3 on Ser 10 with a greater efficiency than IKK- β .

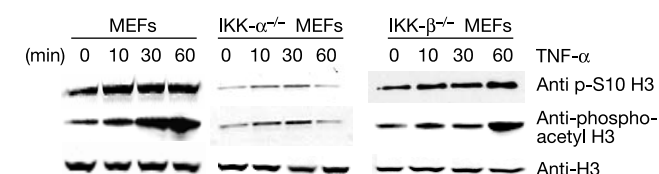


Figure 4 IKK- α modulates global levels of histone H3 phosphorylation. Acid-soluble proteins were extracted from asynchronous IKK wild-type, IKK- α ^{-/-} and IKK- β ^{-/-} MEFs, and were immunoblotted with anti-phospho-specific Ser 10, anti-phosphorylated/acetylated H3, or anti-H3 antibodies.

expression. Current experiments are directed towards investigating the role of promoter-localized IKK- β . Additionally, TNF- α induces recruitment of IKK- γ to the I κ B α promoter, albeit at later time points than IKK- α and IKK- β in wild-type MEFs (data not shown). Collectively, these data support a function for IKK- α that is distinct from the classical role of the cytoplasmic IKK- $\alpha/\beta/\gamma$ complex in controlling cytokine-induced NF- κ B-regulated gene expression. Thus, we propose independent roles for IKK- α , IKK- β and IKK- γ in cytokine-induced NF- κ B-dependent gene expression. Whereas IKK- β controls I κ B α degradation and efficient DNA binding of NF- κ B subunits¹ and IKK- γ acts as a structural or regulatory mediator of this kinase complex^{1,3,21}, IKK- α functions as a chromatin modifier through histone phosphorylation. The data also raise the possibility of a role for IKK- α in NF- κ B-independent gene expression. □

Methods

Cells and reagents

IKK wild-type, IKK- $\alpha^{-/-}$ and IKK- $\beta^{-/-}$ MEFs were provided by I. Verma and M. Karin. Antibodies against IKK- α , IKK- β , IKK- γ , p105/p50, actin and tubulin were obtained from Santa Cruz and Upstate Biotechnology. The p65/RelA-specific antibody was obtained from Rockland. Phosphorylated histone H3 (Ser 10) antibody was obtained from Cell Signaling Technology. Acetylated (Lys 9/Lys 14) histone H3 antibody and phosphorylated/ acetylated (Ser 10/Lys 14) histone H3 antibody was obtained from Upstate Biotechnology. TFIIIB antibody was obtained from Transduction Laboratories. TNF- α (Kamiya) was used at a final concentration of 10 ng ml⁻¹.

Immunofluorescence

IKK wild-type, IKK- $\alpha^{-/-}$ or IKK- $\beta^{-/-}$ cells were seeded onto chamber slides (Nalge Nunc International) and treated with TNF- α (10 ng ml⁻¹) for 15 min. After fixation with 4% paraformaldehyde, cells were permeabilized with 0.2% Triton-X and blocked with 10% goat serum, 1% bovine serum albumin in phosphate-buffered saline (PBS). Primary antibodies were incubated at 4 °C overnight. One-hour incubations with fluorescein isothiocyanate-conjugated secondary antibodies (Santa Cruz Biotechnology) were used to detect primary antibody and were visualized on a Zeiss Axioskop. All images are shown at $\times 200$.

Real-time quantitative PCR

Five micrograms of total RNA was incubated with Moloney murine leukaemia virus-reverse transcriptase (Invitrogen) as recommended by the manufacturer. The resulting complementary DNA was analysed quantitatively for the expression of I κ B α by fluorogenic 5'-nuclease PCR as described previously. Specific primers (forward 5'-AGGATGAGCTGCCCTATGATGA-3' and reverse 5'-TGCCACTTTCCTTATATATGTCAGA-3') and probe (5'-6-FAM TGTGTGTTTGGAGGCCA-TAMRA) were designed to the I κ B α gene and PCR products were continuously measured by means of ABI Prism 7900 during 40 cycles. All data were normalized to 18S ribosomal RNA.

Northern blot analysis

Total RNA was isolated using Trizol (Invitrogen) as recommended by the manufacturer. A total of 10–20 μ g total cellular RNA was separated on 1.5% formaldehyde-agarose gels and transferred overnight to a nylon filter according to standard procedures. RNA was then crosslinked to the membrane by ultraviolet irradiation (Stratagene) and probed with randomly labelled IL-6 probe. Hybridization and wash was performed using ExpressHyb (Stratagene) as described by the manufacturer.

ChIP assay

ChIP analysis was performed following a protocol provided by Upstate Biotechnology under modified conditions. After TNF- α (20 ng ml⁻¹) stimulation, 3×10^6 cells were fixed with 1% formaldehyde. After 5 min, cells were washed extensively with ice-cold PBS and lysed for 10 min in lysis buffer (Upstate Biotechnology). Chromatin was sheared by sonication to an average size of approximately 1 kilobase and pre-cleared for 2 h at 4 °C with salmon sperm DNA-saturated protein G Sepharose. Chromatin solutions were precipitated overnight at 4 °C using 10 μ l anti-p65, 10 μ g anti-IKK- α or anti-IKK- β , 10 μ l anti-phospho-H3 (Ser 10), and 5 μ l acetylated H3-specific antibodies or beads alone. Immune complexes were collected with salmon sperm DNA-saturated protein G Sepharose for 1 h and washed extensively following the manufacturer's protocol. Input and immunoprecipitated chromatin were incubated at 65 °C overnight to reverse crosslinks. After proteinase K digestion, DNA was extracted with phenol/chloroform and precipitated with ethanol. Precipitated DNAs were analysed by PCR (30–35 cycles) using Platinum Taq PCR Master Mix (Invitrogen). The following promoter-specific primers were used: primer pair 5'-TGGCGAGGTCTGACTGTGTGG-3' and 5'-

GCTCATCAAAAGTTCCCTGTGC-3' was used to amplify a 230-base-pair (bp) region in the mouse I κ B α promoter; primer pair 5'-TGTGTGTGTGTGTATGTGTGTGTCG-3' and 5'-TCGTTCTTGGTGGGCTCCAG-3' was used to amplify a 440-bp region in the mouse IL-6 promoter or the β -actin promoter (5'-TGCACTGTGCGCGGAAGC-3' and 5'-TCGAGCCATAAAGGCAA-3').

In vitro kinase assay

Kinase assays were performed following a previously described protocol (Upstate Biotechnology). Kinase activity was determined by incubating purified chicken core histones (10 mg ml⁻¹) with increasing amounts of recombinant IKK- α or IKK- β , or MSK1 (20 mU) and RSK2 (77 U) in the presence of 1 μ Ci ml⁻¹ [γ -³²P]ATP or cold ATP (100 μ M) for 30 min at 30 °C. Reactions were resolved by SDS-polyacrylamide gel electrophoresis (PAGE; 15%) and processed for autoradiography or protein immunoblotting. Recombinant IKK- α and IKK- β were provided by L. Dang.

Western blot analysis

Extractions of acid-soluble proteins were done according to the protocol described by Upstate Biotechnology and resolved on 10–20% Tris-tricine SDS-PAGE gels.

Received 6 March; accepted 28 March 2003; doi:10.1038/01648.

- Ghosh, S. & Karin, M. Missing pieces in the NF- κ B puzzle. *Cell* **109**, S81–S96 (2002).
- Silverman, N. & Maniatis, T. NF- κ B signaling pathways in mammalian and insect innate immunity. *Genes Dev.* **15**, 2321–2342 (2001).
- Courtois, A., Smahi, A. & Israel, A. NEMO/IKK γ : linking NF- κ B to human disease. *Trends Mol. Med.* **7**, 427–430 (2001).
- Senfleben, U. *et al.* Activation by IKK α of a second, evolutionarily conserved, NF- κ B signaling pathway. *Science* **293**, 1495–1499 (2001).
- Li, Q. *et al.* IKK1-deficient mice exhibit abnormal development of skin and skeleton. *Genes Dev.* **13**, 1322–1328 (1999).
- Li, Q. *et al.* Severe liver degeneration in mice lacking the I κ B kinase 2 gene. *Science* **284**, 321–325 (1999).
- Li, X. *et al.* IKK α , IKK β , and NEMO/IKK γ are each required for the NF- κ B mediated inflammatory response program. *J. Biol. Chem.* **277**, 45129–45140 (2002).
- Birbach, A. *et al.* Signaling molecules of the NF- κ B pathway shuttle constitutively between cytoplasm and nucleus. *J. Biol. Chem.* **277**, 10842–10851 (2002).
- Saccani, S., Pantano, S. & Natoli, G. p38-dependent marking of inflammatory genes for increased NF- κ B recruitment. *Nature Immunol.* **3**, 69–75 (2002).
- Ainbinder, E. *et al.* Mechanism of rapid transcriptional induction of TNF α -responsive genes by NF- κ B. *Mol. Cell. Biol.* **22**, 6354–6362 (2002).
- Saccani, S., Pantano, S. & Natoli, G. Two waves of NF- κ B recruitment to target promoters. *J. Exp. Med.* **193**, 1351–1359 (2001).
- Sassone-Corsi, P. *et al.* Requirement of Rsk-2 for epidermal growth factor-activated phosphorylation of histone H3. *Science* **285**, 886–891 (1999).
- Thomson, S. *et al.* The nucleosomal response associated with immediate-early gene induction is mediated via alternative MAP kinase cascades: MSK1 as a potential histone H3/HMG-14 kinase. *EMBO J.* **18**, 4779–4793 (1999).
- Clayton, A., Rose, S., Barratt, M. & Mahadevan, L. Phosphoacetylation of histone H3 on c-fos and c-jun associated nucleosomes upon gene activation. *EMBO J.* **19**, 3714–3726 (2000).
- Thompson, S., Clayton, A. & Mahadevan, L. Independent dynamic regulation of histone phosphorylation and acetylation during immediate early gene induction. *Mol. Cell* **8**, 1231–1241 (2001).
- Crosio, C. *et al.* Mitotic phosphorylation of histone H3: spatio-temporal regulation by mammalian Aurora kinases. *Mol. Cell. Biol.* **22**, 874–885 (2002).
- Berger, S. L. An embarrassment of niches: the many covalent modifications of histones in transcriptional regulation. *Oncogene* **20**, 3007–3013 (2001).
- Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2001).
- Cheung, P. *et al.* Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. *Mol. Cell* **5**, 905–915 (2000).
- Lo, W. *et al.* Phosphorylation of serine 10 in histone H3 is functionally linked *in vitro* and *in vivo* to Gcn5-mediated acetylation at lysine 14. *Mol. Cell* **5**, 917–926 (2000).
- Prajapati, S. & Gaynor, R. Regulation of I κ B kinase (IKK γ /NEMO) function by IKK β -mediated phosphorylation. *J. Biol. Chem.* **277**, 24331–24339 (2002).
- Yamamoto, Y., Verma, U. N., Prajapati, S., Kwak, Y.-T. & Gaynor, R. B. Histone H3 phosphorylation is critical for cytokine-induced gene expression. *Nature* **423**, 655–659 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D. Kashatus for critical reading of the manuscript. We appreciate scientific discussions with Y. Yamamoto and R. Gaynor. Support was provided by the NIH and by the UNC breast cancer SPOR program to A.S.B., and by the UNC Comprehensive Center for Inflammatory Disorders.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.S.B. (jhall@med.unc.edu).

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/

Iberia/ Belgium:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria:

Patrick Phelan, Odo Wulffen

Tel +49 89 54 90 57 11/-2

Fax +49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Haraikatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Hideki Watanabe

e-mail: h.watanabe@naturejpn.com

naturejobs

Food for thought

Some bench scientists might turn their noses up at a career in the food industry — not enough room for problem-solving and pure research, they might say. But they'd be wrong, two industry researchers told an alternative careers session at the American Society for Microbiology meeting in Washington last month.

Tom Pugh, director of applied research and development at E&J Gallo Winery in Sonoma Valley, California, noted that even though wine-making relies primarily on one chemical process — fermentation — many microbiological nuances can affect it. Identifying and managing myriad bacteria and fungi that can damage a vintage is a key role for microbiologists, he told the meeting. Typical quality-control work tends to be done by scientists with bachelor's or master's degrees, he said, whereas more basic research, such as understanding the impact that different microorganisms have on fermentation, tends to be done by PhDs.

Kayla Polzin, who works for the Minneapolis dairy company Land O'Lakes, said that being a scientist in the dairy industry is akin to being a detective. A frequent question she and her colleagues ask is: "What is the microorganism and where did it come from?" She recalled that one of her first duties at the company was to try to determine why one vat of parmesan cheese tasted bland whereas another tasted nutty — even though the two had come from the same batch of milk. She sees growth potential for careers in dairy, especially in creating 'functional foods', which have bacteria added to them to confer various health effects.

Qualifications for work in wine and cheese require a good background in microbiology. And both industries emphasize an ability to solve problems. But these skills are not dissimilar to their bench counterparts — making snobbery in either direction unnecessary.

Paul Smaglik

Naturejobs editor



Contents

SPECIAL REPORT

Swapping the lab
for the law

p666

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS



For scientists who want a career in law, the options are greater than ever – and some don't even involve going back to school, says Amy Wilson.

As a lawyer with a background in science, Jennifer Gordon has skills that allow her to switch from the courtroom to the laboratory bench without any problems. Gordon was among the first wave of life-science PhDs who joined firms that concentrate on patent law. In the early 1980s, while she was doing her PhD in biochemical engineering at the Massachusetts Institute of Technology in Cambridge, she met Leslie Misrock, a partner at Pennie & Edmonds in New York, who wanted to build a law practice that would focus on biotechnology, a field that was then in its infancy.

Gordon joined the firm as a law clerk in 1981, working in the office during the day and attending Fordham University School of Law at night. The firm's clerk programme covered the cost of law school and allowed her to gain valuable work experience.

Now a partner at the New York office of law firm Orrick, Herrington & Sutcliffe, Gordon has watched patent law relating to biotechnology develop. When she started out, there was no biotechnology-specific case law or legislation; now, US patents cover everything from single genes to multicellular organisms. Looking back, the opportunities for growth seemed obvious. "It became clear that we were in an area that needed its own law," Gordon says.

As the law has evolved, so have the options for graduates wanting to pursue a career in science-related law. Clerk programmes that sponsor a PhD through law school now exist both in the United States and across Europe. But there are also opportunities for scientific PhDs who don't want to do any formal education at law school — they can become registered patent agents or technology-transfer specialists.

To become a US patent agent, scientific graduates must pass the exam set by the US Patent and Trademark Office (PTO). If successful, they are allowed to write patents and represent clients before the PTO. Law firms, universities and pharmaceutical and biotechnology companies routinely employ patent agents.

Michael Yamin, a registered patent agent, took an entrepreneurial approach to his career. After his academic training, stints with two biotech companies and the non-profit Picower Institute for Medical Research in Manhasset, New York, exposed him to various aspects of patent law. As he enjoyed those encounters, he decided to join a law firm. Several years

later, Mojave Therapeutics, a biotech company in Hawthorne, New York, recruited him as director of intellectual property and licensing.

At Mojave he manages the company's patent portfolio and licensing agreements. He advises graduates who want to pursue a career in a legal field to manage their expectations. Patent agents tend to have a less clearly defined role inside a company than patent lawyers, and need to have realistic expectations of their duties. For example, a patent agent without a law degree would not be allowed to handle litigation.

BROAD APPLICATION

Becoming a patent agent in Europe also involves sitting a qualifying examination. But the exam covers more ground than its US equivalent. Passing it confers the rank of European patent attorney, rather than agent, which allows you to take on additional responsibilities such as writing opinions on patent infringements.

Andrea Carreira-Staubli is well aware of the differences. She is a patent agent in the United States, works at a law firm in Switzerland, and is studying for the European qualifying examination. She was born in Switzerland, where she received her masters in chemistry. After earning her PhD in applied biological sciences at the Massachusetts Institute of Technology, she did two postdocs at the California Institute of Technology in Pasadena, where she realized that she wanted to be involved in science, but didn't want to stay at the bench.

As she enjoyed her work preparing patent applications during her postdoc, Carreira-Staubli decided to take the patent agent's exam in the final year of her fellowship. After certification, she worked at a Los Angeles law firm with the intention of eventually attending law school at night.

But after one year she moved back to Switzerland, where she worked as a trainee patent attorney for three years with drug firm Novartis in Basel before joining Isler & Pedrazzini, a Zurich law firm, in 2001. She is responsible for all of the firm's biotechnological and chemical patent applications.

Carreira-Staubli describes the work in Europe as less hectic than in the United States. The hours involved allow the attorneys to invest the appropriate time in client care, as well as leaving time for

From entrepreneur to patent agent: Michael Yamin.



Warning: this is a copyright law and international treaties

continuing their personal education, she says.

For graduates who want to avoid both law school and patent-agent exams, the role of technology-transfer specialist offers another route into law. This kind of work provides a link between intellectual property and business for companies or universities, and the individuals involved tend to be self-taught — and self-motivated.

That profile fits Jahanara Ali, associate director of business development at Mojave Therapeutics. Ali obtained a PhD in molecular biology from New York University and then did two postdocs, one at

Rockefeller University and the other at the Cleveland Clinic in Ohio. While in Cleveland, she contacted the director of the technology-transfer office at New York University and asked about the type of work performed there. He gave her a market-research project on different inventions coming out of the university, and she was hooked. Through this work, she landed a position at Mount Sinai School of Medicine's technology-transfer office in New York.

Her responsibilities at Mount Sinai included ensuring patent protection for scientist inventors, and licensing inventions to biotech and pharmaceutical companies, in addition to reviewing and drafting material-transfer agreements, sponsored research and collaboration agreements.

Ali believes that having a broad legal experience has helped her current role in business development, which combines reviewing legal agreements, identifying partnering opportunities, and monitoring activities of the company's key competitors. She and Yamin work closely together and agree that business development and intellectual property are intimately related, especially in a small company.

Careers in law tend to remain flexible, as Hayley French's experiences illustrate. After she finished her microbiology PhD in London, she decided not to do a postdoc. "I just knew I didn't want to be at the bench," she says. Instead she turned to the commercial-law department and took a master's degree in intellectual property law. That led to jobs in intellectual property management, first with the agriculture ministry, then with University College London. But she eventually decided to become a fully fledged lawyer, meaning that she had to take a conversion course (see 'English options'; left) and complete a two-year apprenticeship.

Although career paths in law for scientists are now more varied than they were 20 years ago, they still share a common origin. Scientists who choose law careers — whether as lawyers, patent agents or technology-transfer specialists — first establish a foundation in science, then apply their technical expertise to the law. Those who choose law-related careers emphasize that they have not abandoned their roots. Instead, they say that they have found a way to learn about new fields and technologies. And what true scientist wouldn't like that?

Amy Wilson is a technical consultant for law firm Stroock & Stroock & Lavan in New York.

English options

As a UK law degree is an undergraduate qualification, becoming a lawyer in Britain after earning a scientific PhD is a slightly different process from the clerk programmes in the United States. Scientists take a conversion course, which covers the undergraduate law curriculum in an accelerated and intensive format, followed by an additional postgraduate diploma and a two-year apprenticeship, required for all UK lawyers.

Myles Jelf took that route after he finished his PhD in carbon-fibre composites at the University of Cambridge. During his postdoc in the engineering department at Cambridge, Jelf came across a



Myles Jelf enjoys merging technology and law.

presentation by Bristows, a technology-focused law firm that actively recruits PhD scientists.

Bristows gave Jelf a job at its London office, where he focuses broadly on patent litigations and other areas of intellectual property such as trademarks and copyrights. He says that he particularly enjoys integrating his knowledge of scientific technologies with his understanding of the law.

A.W.

Trail-blazer: Jennifer Gordon was one of the first scientific PhDs to move into biotech law.

